

Máster Interuniversitario en Estadística e Investigación Operativa UPC-UB

Título: Caracterización y repetibilidad de la mecánica de las trayectorias de vuelo del *Larus Audouinii*

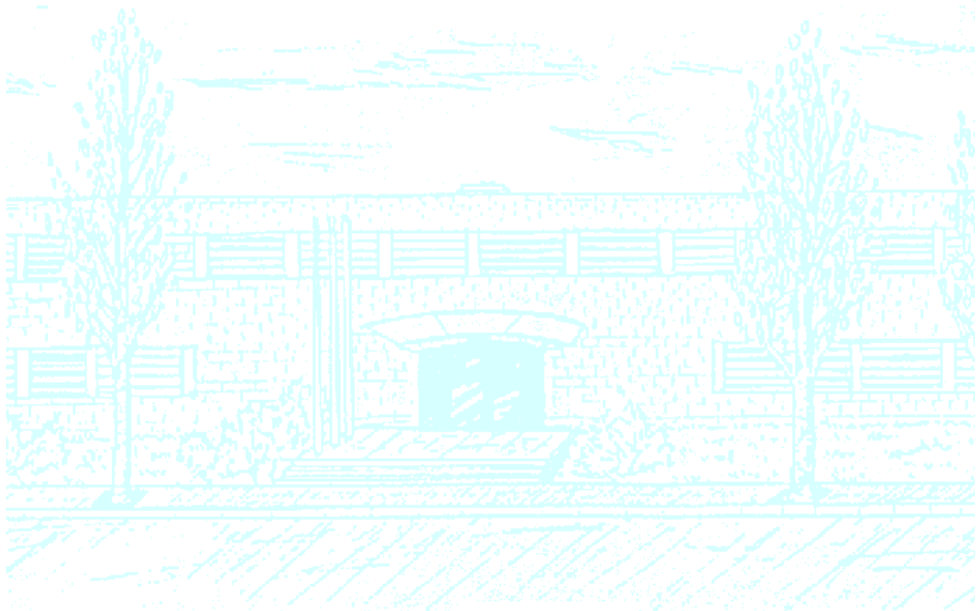
Autor: Augusto Cortejana Retamozo

Director: Josep Lluís Carrasco

Departamento: Fundamentos Clínicos

Universidad: Universidad de Barcelona

Convocatoria: Enero 2020



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística



UNIVERSITAT DE BARCELONA

Facultad de Matemáticas y Estadística
Universidad Politécnica de Cataluña

Máster en Estadística e Investigación Operativa (MESIO)
Trabajo Final de Máster

**Caracterización y repetibilidad de la
mecánica de las trayectorias de vuelo
del *Larus Audouinii***

Augusto Cortejana Retamozo

Director: Josep Lluís Carrasco

Departamento de Fundamentos Clínicos

Agradezco de corazón a todas las personas que me han acompañado en esta etapa universitaria: a mi familia por enseñarme buenos valores y a mis amigos por los buenos momentos. Finalmente, gracias a mi tutor de trabajo final de máster, Josep Lluís Carrasco por guiarme en la realización del presente trabajo.

Resumen

En los últimos años, ha crecido la disponibilidad de datos georeferenciados para estudiar el comportamiento de los animales: composición de su hábitat, estudios de migración, entre otros. La trayectoria de los animales es una parte muy importante de estos estudios: por ejemplo, cuanto de tortuosa es una trayectoria o las distintas velocidades que presenta proporciona información sobre el comportamiento de los animales.

Este trabajo se centra en estudiar la mecánica de las trayectorias de vuelo de las gaviotas *Larus Audouinii* del Delta del Ebro a partir de la construcción de índices que caracterizan estas trayectorias como la sinuosidad, los cambios direccionales, velocidad y aceleración.

Mediante los modelos lineales mixtos, se analizará cada índice para ver si hay diferencias significativas según el tipo de día y el momento de inicio de la trayectoria. Se utiliza esta clase de modelos ya que los datos de las trayectorias tienen un efecto aleatorio clúster que es la gaviota. Además, se evalúa la repetibilidad del índice mediante el coeficiente de correlación intraclases. Para ver la relación entre los índices y encontrar las variables latentes que caracterizan las trayectorias se utiliza el Análisis de Componentes Principales.

Finalmente se construye un aplicativo Shiny para la visualización y comparación de trayectorias utilizando leaflet, una librería JavaScript para crear mapas interactivos.

Palabras claves: *trayectoria de animales, software R, modelos lineales mixtos, PCA, repetibilidad, Larus Audouinii, Shiny, leaflet*

MSC2000: 2000Mathematical Subject Classification

Abstract

In recent years, the availability of georeferenced data to study the behavior of animals has increased: composition of their habitat, migration studies, among others. The trajectory of animals is a very important subject of animal behavior studies: for example, how tortuous a trajectory is or the different speeds it presents provides information about animal behavior.

This work focuses on studying the mechanics of the flight paths of the *Larus Audouinii* gulls at the Delta del Ebro through the construction of indexes that characterize these trajectories such as sinuosity, directional changes, speed and acceleration.

Using the mixed linear models, each index will be analyzed to see if there are significant differences according to the type of day (workdays and weekdays) and the moment of beginning of the trajectory (day, afternoon and night). This type of models is used because the trajectory data has a random effect that is the seagull. In addition, the repeatability of the index is assessed using the intraclass correlation coefficient. The Principal Component Analysis is used in order to analyze the relationship between the indexes and find the latent features that characterize the trajectories

Finally, a Shiny application for the visualization and comparison of trajectories is built using leaflet, a JavaScript library to create interactive maps.

Key words: *animal's trajectory, R software, mixed linear models, PCA, repeatability, Larus Audouinii, Shiny, leaflet*

MSC2000: 2000Mathematical Subject Classification

ÍNDICE

Introducción	1
Motivación y objetivos	1
Estructura del trabajo	2
CAPÍTULO 1: Metodología	3
1.1 Definición de trayectoria	3
1.2 Índices para caracterizar las trayectorias	3
1.3 Modelos lineales mixtos	6
1.4 Índice de repetibilidad (ICC)	7
1.5 Análisis de Componentes Principales	8
1.6 Paquete trajr	9
1.7 Shiny	9
CAPÍTULO 2: Datos del análisis	11
2.1 Presentación de los datos	11
2.2 Preproceso de los datos	13
2.3 Descriptiva	14
2.4 Obtención de los datos a nivel trayectoria	16
CAPÍTULO 3: Modelos mixtos para el análisis de los índices	17
3.1 Análisis de la sinuosidad	17
3.2 Análisis de la duración	22
3.3 Análisis de la no linealidad (media del cambio direccional)	27
3.4 Análisis de la irregularidad (desviación típica del cambio direccional)	32
3.5 Análisis de la velocidad media	38
3.6 Análisis de las desviaciones de velocidad	43
3.7 Análisis de la aceleración media	48
3.8 Análisis de las desviaciones de aceleración	53
CAPÍTULO 4: Análisis de Componentes Principales	57
4.1 Representación de las variables numéricas	58
4.2 Representación conjunta de variables en el plano factorial	59
4.3 Interpretación de las dos primeras componentes	60
CAPÍTULO 5: Aplicativo Shiny	61
Conclusiones	64

Introducción

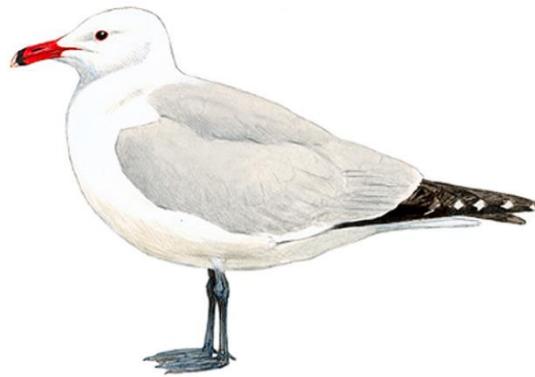
En los últimos años, ha crecido la disponibilidad de datos de rastreo de animales la cual cosa brinda nuevas oportunidades para intentar entender más su comportamiento a través de técnicas estadísticas.

Este trabajo se centra en analizar las trayectorias de vuelo de las gaviotas de Audouin (*Larus audouinii*) situadas en la colonia de Punta de la Banya en el Delta del Ebro.

La gaviota de Audouin es muy escasa y rara. En España, en los años 70 había apenas unos cientos de parejas. Fue en la siguiente década donde comenzó a colonizar otros enclaves litorales como el del delta del Ebro y empezó a crecer bastante en número de efectivos a diferencia de otros sitios del mundo en donde desaparecieron importantes colonias de cría.

Es por ello por lo que, en la actualidad, se considera a la población española de gaviotas Audouin como la más importante a nivel mundial (acoge al 90% de los efectivos mundiales) y, el delta del Ebro la zona de cría más importante de esta especie.

La especie se encuentra incluida en el Libro Rojo de las aves de España en la categoría de “Vulnerable” y aparece como “De interés especial” en el Catálogo Nacional de Especies Amenazadas. (Fuente: SEO BirdLife)



Motivación y objetivos

El análisis de trayectorias es una temática compleja y novedosa. Se trata de un análisis espacial particular y del que no existe una gran cantidad de literatura en la actualidad. Además, el análisis de repetibilidad propuesto implica que, en su conjunto, se deban aplicar técnicas como los

modelos mixtos, el PCA y test de permutaciones. Lo que supone un reto y una oportunidad de aplicar y desarrollar los conocimientos adquiridos en el máster.

El presente trabajo se centrará en la consecución de los siguientes objetivos principales:

1. Describir la mecánica del vuelo de las gaviotas mediante diferentes índices
2. Evaluar la repetibilidad de estos índices
3. Investigar si las características de las trayectorias cambian según el tipo de día de salida y el momento del día

Estructura del trabajo

El trabajo se estructura de la siguiente manera:

En el capítulo 1 se explica la metodología usada: índices de trayectoria, modelos mixtos, el índice de repetibilidad, el análisis de componentes principales y el paquete `trajr` de R.

En el capítulo 2, se describe el origen, recogida y manipulación de los datos. La fase de preproceso que incluye la creación y eliminación de variables, la corrección de errores y la construcción de los índices a nivel de cada trayectoria.

Después, y una vez obtenido el diseño final de los datos (una fila por trayectoria y cada columna es un índice), en el capítulo 3 se analiza cada índice a partir de modelos mixtos y comparación de modelos y se evalúa la repetibilidad del índice a partir del coeficiente de correlación intraclases.

Posteriormente, en el capítulo 4 se realiza un análisis multivariante a partir del Análisis de Componentes Principales con el objetivo de ver la relación entre índices e intentar encontrar las variables latentes que caracterizan las trayectorias.

Finalmente, el último capítulo explica la construcción de un dashboard en Shiny para la visualización de todas las trayectorias.

CAPÍTULO 1: Metodología

En este capítulo inicial se explicarán las diferentes técnicas estadísticas e índices utilizados a lo largo del trabajo para analizar las trayectorias.

1.1 Definición de trayectoria

Una trayectoria es la simplificación de un camino real realizado por un animal. También es un conjunto de coordenadas espaciales bidimensionales junto con la tercera dimensión que es la temporal. Por último, también se puede definir una trayectoria como una serie de pasos, cada uno con una longitud, un turning angle (Δ_i) y una duración.

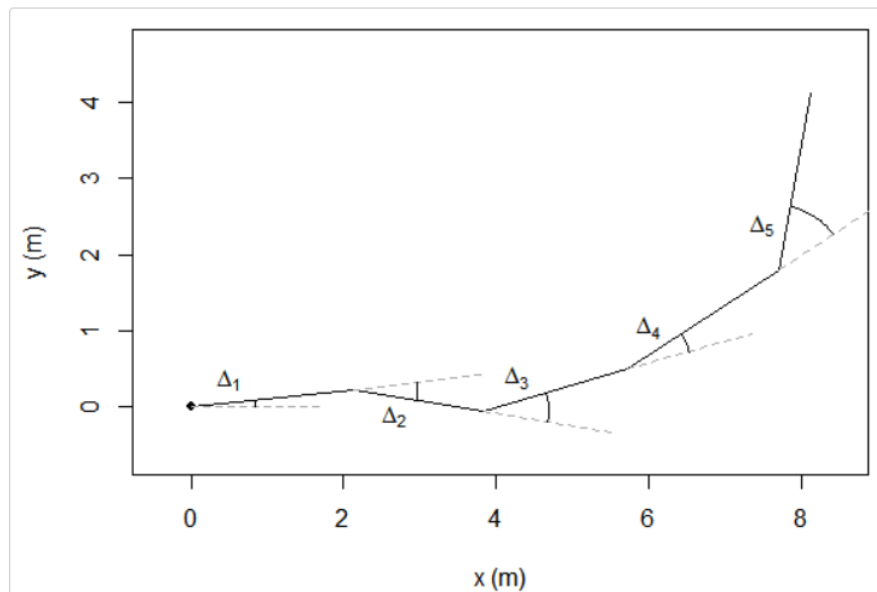


Figura 1.1: Definición de trayectoria

1.2 Índices para caracterizar las trayectorias.

A continuación, se explica en detalle los índices empleados para describir la mecánica del vuelo de las gaviotas.

Índices de velocidad y aceleración

Se define la velocidad como el desplazamiento recorrido por unidad de tiempo y la aceleración lineal como el cambio en la velocidad entre el tiempo.

Hay diferentes valores de velocidad y aceleración a lo largo de una misma trayectoria por lo que lo interesante de estos índices es que los cambios de velocidad sirven para identificar diferentes tipos de comportamiento en animales: por ejemplo, se sabe que la gaviota está descansando si la velocidad es <1 km/h o en movimiento en caso contrario.

Se utilizará como índices la media y desviación típica de las velocidades y aceleraciones de cada trayectoria.

Para estas dos medidas, hace falta primero de todo reducir el ruido en la trayectoria ya que contamina la calidad de la medición. Para ello, es común suavizar la trayectoria mediante el filtro de Savitzky-Goley (Jim McLean 2018).

Este filtro consiste en aplicar una regresión polinómica local de grado k con al menos $k+1$ puntos equiespaciados para determinar el valor suavizado de cada punto original.

Índices de tortuosidad

El nivel de tortuosidad (también llamado sinuosidad) es un factor clave para caracterizar una trayectoria. Se define como el nivel de curvas de una trayectoria. El concepto opuesto a la tortuosidad es el nivel de rectitud.

Existen diferentes índices para medir la tortuosidad: el índice de rectitud d que es el más simple, el índice de sinuosidad S corregida, la dimensión fractal que es de las más utilizadas y el índice de desplazamiento máximo esperado.

Pero algunos presentan inconvenientes, por ejemplo, el índice d es correcto solo para caminos no aleatorios, la dimensión fractal no es apropiada para trayectorias de animales ya que no suelen caracterizarse por ser curvas fractales (Nams, 2006; Turchin, 1996).

Para este análisis se ha decidido utilizar el índice de sinuosidad corregido:

Este índice es una versión corregida del índice S de sinuosidad por Benhamou en 2004 que es más generalizable para un mayor rango de variaciones en “step angles”

$$S_{corr} = \frac{2}{\sqrt{p \left(\frac{1+c}{1-c} + b^2 \right)}}$$

donde c es el coseno medio de los *turning angles*, b es el coeficiente de variación de longitud de paso y p es la longitud media de paso.

Si la trayectoria tiene una longitud de paso variable, se aconseja rediscretizarla a una longitud de paso constante antes de calcular este índice.

Dos ejemplos de trayectoria: en la Figura 1.2, la de la izquierda presentaría un índice de sinuosidad bajo mientras que la de la derecha un valor alto al realizar un camino más tortuoso.

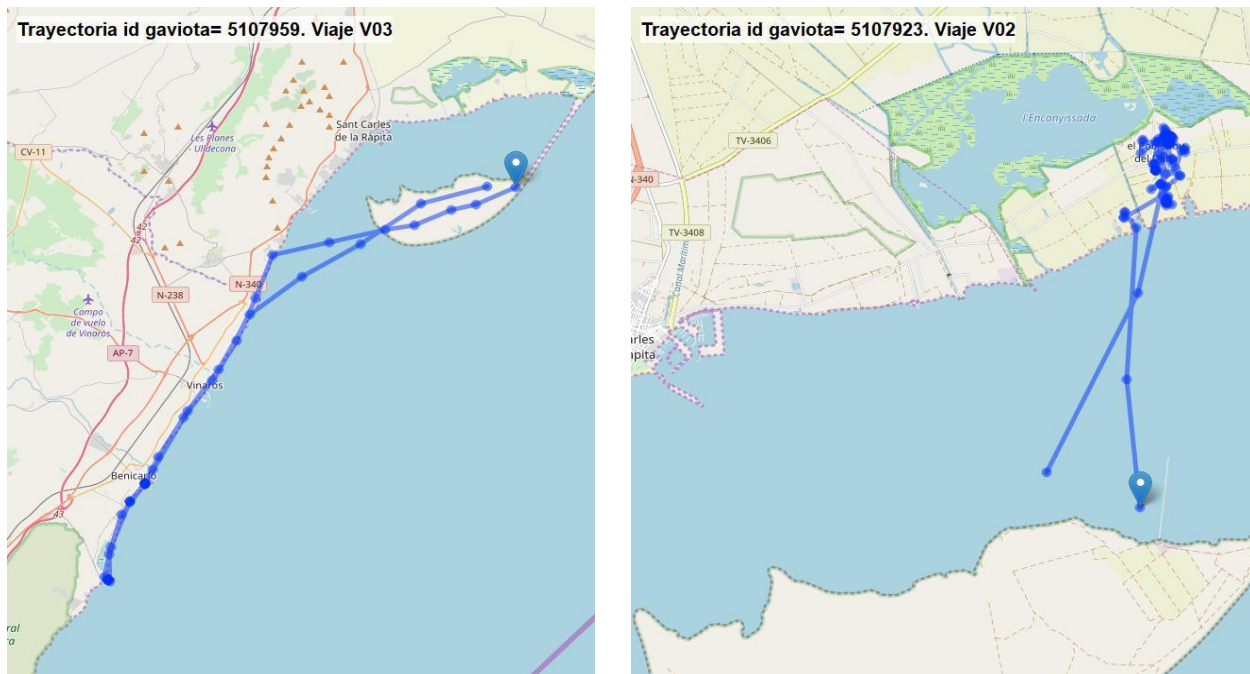


Figura 1.2: Ejemplo de trayectorias para comparar sinuosidad (izq: baja sinuosidad, dcha: alta sinuosidad)

Cabe remarcar la diferencia entre trayectorias aleatorias y orientadas ya que la tortuosidad de una trayectoria se puede medir ya sea con un índice de rectitud o con uno de sinuosidad dependiendo si el animal realizó un movimiento orientado o uno de aleatorio.

Por ejemplo, en experimentos de búsqueda, los animales se orientan hacia su hogar con una mayor o menor eficiencia. En cambio, en otros experimentos, se asume que van libremente por el entorno, por ejemplo, buscando objetos que no se pueden detectar a distancia.

Pero hay experimentos en que esta clasificación no es del todo clara: por ejemplo, el animal alcanzó su objetivo realizando un camino de búsqueda aleatorio o puede haberse orientado activamente hacia el objetivo.

Los investigadores aseguran que un camino muy recto sugiere un movimiento más orientado, en cambio una mayor tortuosidad en el camino puede ser debida de una orientación menos eficiente, así como de una búsqueda aleatoria de baja sinuosidad (Benhamou, 2004).

Esto se puede observar en las dos trayectorias de ejemplo: la de la derecha que es más tortuosa se trata de una gaviota volando en un arrozal que se está alimentando. La otra trayectoria sigue la línea de la costa hasta el castillo de Peñíscola y vuelve al nido.

Índice de cambio direccional DC

El índice de cambio direccional mide el cambio en la dirección de la trayectoria en el tiempo (Kitamura & Imafuku, 2015).

La fórmula es la siguiente:

$$DC_{v_1, v_2} = \frac{\left(\frac{180}{\pi} \arccos \left(\frac{v_1 v_2}{|v_1| |v_2|} \right) \right)}{t}$$

El cambio direccional, por defecto, se calcula para cada par de posiciones consecutivas v_1 y v_2 : se calcula el ángulo entre dos vectores de posición y se transforma en radianes. Por último, se divide entre el tiempo que tarda entre las dos posiciones.

A diferencia de los índices de tortuosidad y de rectitud, este índice incorpora la velocidad del cambio, es decir, indica la rapidez y frecuencia en que un animal cambia su dirección.

La media del índice se utiliza como medida de no linealidad y la desviación típica como medida de irregularidad.

1.3 Modelos lineales mixtos

Cuando los datos se encuentran agrupados con una estructura de clúster la premisa de independencia de las observaciones que se utiliza en los modelos lineales no se cumple. Los modelos lineales mixtos son una de las soluciones más utilizadas para incorporar este tipo de dependencia de los datos.

Los modelos lineales mixtos son modelos de regresión lineal que incorporan dos tipos de efectos:

- Efectos fijos. Explican variabilidad global de la respuesta en función de los valores de las covariables (cuantitativas y cualitativas). Sus valores son comunes para aquellos individuos con la misma configuración de covariables.
- Efectos aleatorios. Sus valores varían entre clústers (individuos). Por lo tanto, explican variabilidad que no es recogida por los efectos fijos.

Esta combinación de diferentes efectos (fijos y aleatorios) justifica el nombre de modelos mixtos. Paramétricamente el modelo se expresa como:

$$Y_i = X_i \beta + Z_i b_i + \varepsilon_i \quad b_i \sim N(0, D), \quad \varepsilon_i \sim N(0, \sigma^2 I)$$

Donde Y es el vector de respuestas, X es la matriz de covariables implicadas en los efectos fijos, Z es la matriz asociada a los efectos aleatorios b que son normales con matriz de varianzas y covarianzas D

Estimación de los modelos mixtos

Los parámetros del modelo pueden ser estimados mediante máxima verosimilitud (ML), pero en este caso los estimadores de los componentes de la varianza están sesgados (la esperanza del estimador infraestima los parámetros). La solución es la estimación vía Verosimilitud Restringida (REML) que estima primero los efectos fijos y luego en el espacio residual, las componentes de la varianza.

En ML se encuentra el valor de los parámetros que maximizan el log verosimilitud estimando a la vez los efectos fijos y aleatorios. El uso de esta estimación es correcto al comparar modelos con diferentes efectos fijos o al comparar un modelo mixto con un modelo lineal.

En REML, lo que interesa es estimar los efectos aleatorios. Restringe el espacio de parámetros limitando el valor de los efectos fijos y en este espacio restringido, busca los valores de los efectos aleatorios que maximizan la log verosimilitud. Se repite este proceso varias veces y por último da la media como estimaciones finales.

En nuestro caso, los datos no son independientes: el **efecto aleatorio clúster es la gaviota**, y los **efectos fijos** serán **el tipo de día** de inicio de la trayectoria (día laborable o fin de semana) **y el momento del día** en que se inicia la trayectoria (mañana, tarde o noche). La interacción entre estos dos factores también se tendrá en cuenta.

Para comparar diferentes modelos se utilizará el test de razón de verosimilitudes (en adelante LRT) y el criterio de información de Akaike (AIC).

El estadístico del LRT test es el siguiente:

$$LRT = -2 \ln \frac{L_1}{L_2} \sim \chi_{df2-df1}^2$$

Donde L_1 es la verosimilitud del modelo 1 y L_2 la del modelo 2 siendo este último el más complejo. Si el p-valor del test es menor que 0.05 se considerará que el modelo 2 es preferible al modelo 1.

La fórmula del criterio AIC es:

$$AIC = 2k - 2\ln(L)$$

Donde k es el número de parámetros del modelo. La primera parte de la ecuación es una penalización a la log verosimilitud por la complejidad del modelo (más número de parámetros, más complejo y por lo tanto más penalización). La regla de decisión es escoger el modelo con menor AIC.

En el presente trabajo se utilizarán estos dos criterios para la comparación de modelos con diferente estructura de varianzas y covarianzas y así elegir la mejor estructura para cada índice. Una vez elegida la mejor estructura de varianzas, se compara modelos con diferentes efectos fijos (tipo de día, momento del día e interacción).

1.4 Índice de repetibilidad (ICC)

La repetibilidad es una medida de la consistencia entre medidas repetidas de una misma característica en un mismo individuo, es decir, el grado de concordancia entre observadores al medir una característica continua. También se le conoce como coeficiente de correlación intraclase.

Con los modelos mixtos, el coeficiente de correlación intraclase es la ratio de la varianza entre clústers (varianza entre gaviotas) entre la varianza total de la variable respuesta.

El índice de repetibilidad, por lo tanto, está acotado entre 0 y 1:

- Si $ICC = 0$, indicaría que las medidas repetidas son tan distintas como si se hubieran tomado a partir de individuos distintos tomados al azar. En este análisis, indica que hay máxima aleatoriedad y el modelo mixto no era necesario.
- Si $ICC = 1$, la medida es consistente y repetible y no se comete ningún error en la característica medida. En este estudio indicaría máxima separación de las gaviotas respecto un determinado índice, es decir, hay un comportamiento totalmente diferente entre gaviotas.

En el trabajo se calculará el ICC de cada índice para ver su repetibilidad, así como también se calculará y segregando por las trayectorias de días laborables y las de fin de semana.

Para ver si hay diferencias significativas en cuanto a la repetibilidad entre días laborables y fin de semana, se utiliza el test de permutaciones:

Permutando la etiqueta de tipo de día (pero respetando la estructura de clúster de los datos) se realizan 1000 permutaciones de datos y en cada una se calcula la diferencia de ICC para días laborables e ICC para el fin de semana.

Por último, se calcula como p-valor, el número de veces en que las diferencias con los datos permutados son mayores que la diferencia de ICCs con los datos originales (ambas diferencias en valor absoluto ya que se trata de un test de hipótesis bilateral) dividido entre el total de permutaciones. Si el p-valor es >0.05 indicara que no hay diferencias estadísticamente significativas.

1.5 Análisis de Componentes Principales

El Análisis de Componentes Principales (en adelante, PCA) es una técnica de reducción lineal de la dimensionalidad de la matriz de datos X .

El objetivo del PCA es reducir la dimensión de un conjunto de variables a un conjunto más pequeño de variables no observadas, llamadas componentes principales o variables latentes.

El criterio para obtener las componentes principales es maximizar la variabilidad explicada de la nube de puntos. es decir, el objetivo es encontrar el mejor plano, que es aquel que al proyectar en él los datos la inercia proyectada sea máxima.

Primero de todo busca la combinación lineal de variables que maximice la inercia total. Esta es la primera componente principal y es la que explica el mayor % de variabilidad. Las demás componentes se calculan teniendo en cuenta:

- Cada componente ha de ser ortogonal (incorrelacionada) con la anterior
- Cada componente tiene que explicar el máximo de variabilidad

Existen tantas componentes principales como número de variables numéricas haya en la matriz de datos. Los valores de éstas se llaman scores y son las proyecciones de los individuos en las componentes principales. Estos valores se obtienen a partir de la descomposición de valores singulares de la matriz de datos X .

En el trabajo se realiza un PCA de los índices de trayectorias para ver cómo se relacionan entre sí e intentar interpretar las variables latentes asociadas a explicar las trayectorias.

1.6 Paquete trajr

Trajr es el paquete de R que se ha utilizado para el análisis y caracterización de trayectorias de animales. Trajr opera con las trayectorias a través de coordenadas espaciales en dos dimensiones y una tercera dimensión que es la temporal. (Mc Lean, 2019)

Para crear una trayectoria a partir de unos datos, se utiliza la función `TrajFromCoords` que requiere como input un dataframe que contenga las coordenadas cartesianas bidimensionales y opcionalmente los tiempos para los puntos de la trayectoria.

El resultado final, es un objeto con clase “trajectory” con 6 columnas:

- `x`: Coordenada X de cada punto
- `y`: Coordenada Y de cada punto
- `time`: el tiempo para cada punto
- `displacementTime`: los tiempos relativos de cada punto.
- `polar`: coordenadas representadas como números complejos para simplificar el análisis con los ángulos.
- `Displacement`: Son los vectores de desplazamiento representados como números complejos entre cada par de puntos consecutivos

Esta es la única información necesaria para calcular los índices para caracterizar trayectorias.

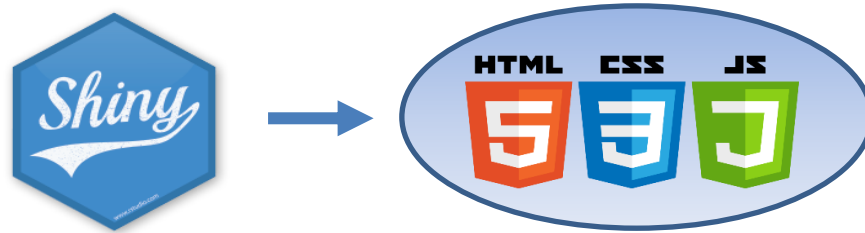
También ofrece las herramientas necesarias para modificar las trayectorias antes del análisis: escalar las trayectorias, suavizarlas para reducir ruido, rediscretizarlas (remuestrear una trayectoria para que tenga una longitud de paso constante; condición necesaria para la construcción de algunos índices, rotarlas).

Además, también se puede visualizar cada trayectoria, así como crear otras completamente aleatorias.

1.7 Shiny

Para visualizar información que proporcionan las trayectorias de las gaviotas de este estudio, se decidió utilizar la herramienta Shiny.

Shiny es un paquete de R creado en el año 2012 por Rstudio. Su función principal es crear aplicaciones interactivas mediante código R que luego el mismo paquete traduce a lenguaje HTML, CSS y javascript logrando así una aplicación web.



La aplicación siempre se compone de dos partes:

- User-interface script (UI): contiene el diseño y el aspecto de la aplicación como por ejemplo cuantas pestañas tendrá la aplicación, donde se ubicará cada elemento a visualizar, etc.
- Server script (SERVER): A partir de los objetos definidos en el UI se construye los resultados a mostrar.

Los objetos de una aplicación son de dos tipos:

- Inputs: conjuntos de valores que el usuario puede introducir y cambiar mediante los widgets (cajas de selección, botones, inputs numéricos, etc). Gracias a ellos, el usuario interacciona con la aplicación. Los inputs se construyen en el UI.
- Outputs: Son los objetos que cambian en función de los inputs mediante reactividad. Pueden ser tablas, gráficos, textos, imágenes y un largo etc. La creación de estos se realiza en el server.

CAPÍTULO 2: Datos del análisis

2.1 Presentación de los datos

El área de estudio se define a partir de los movimientos de las 36 gaviotas Audouin rastreadas por GPS en la colonia de Punta de la Banya, una península arenosa protegida con salinas en el Parque Natural del Delta del Ebro.



Figura 2.1: Mapa de la zona de estudio

El estudio de estas gaviotas se inició durante el período de incubación del año 2011 y en ese momento había alrededor de 10.000 parejas reproductoras, aproximadamente el 68% de la población mundial de gaviotas Audouin (Oro et al., 2010).

Se capturaron 60 gaviotas en nidos escogidos aleatoriamente a partir de trampas etiquetadas en cajas o en carpas y equipadas con el registrador GPS CatTrack (Perthold 2011). Estos GPS registran la ubicación de la gaviota con una precisión de 10 metros cada 5 minutos y se les colocó en la parte posterior de cada gaviota (el dispositivo estaba sellado con un tubo de goma para garantizar su impermeabilidad). Además, el peso del dispositivo era de 25 gramos y no

provocaba efectos nocivos en las gaviotas al estar el peso por debajo del límite (3-5% de la masa corporal de la gaviota) (Wilson et al., 2002; Phillips et al., 2003).

Finalmente, después de 2 semanas desde la colocación de los dispositivos, se capturaron un total de 36 gaviotas marcadas. En el período temporal, que va del 8 al 26 de mayo, no hubo ninguna condición meteorológica adversa como vientos fuertes o lluvia que pudiera sesgar su comportamiento alimentario.

El dataframe tiene **38090 registros** y **17 variables**. Para cada fila del dataframe a analizar se tiene información geolocalizada de la trayectoria de una gaviota en un determinado momento: coordenadas de longitud y latitud, distancia a su nido, velocidad de vuelo y en qué clase de lugar vuela. En conclusión, se entiende cada fila como un punto de la trayectoria de un determinado viaje.

A continuación, se explica brevemente el tipo y significado de las variables.

Nombre variable	Tipo de variable	Significado
<i>Campanya</i>	Único valor	Campaña del análisis: 2011-05 Delta
<i>ANILLA</i>	Numérica	Identificador de la gaviota.
<i>idViaje1</i>	Categórica	Identificador del viaje
<i>idViaje2</i>	Categórica	Identificador trayectoria: id ANILLA +idViaje
<i>fechaOfici</i>	Fecha	Fecha registro de la posición de la gaviota
<i>HoraOficia</i>	Hora	Hora registro de la posición de la gaviota
<i>intervHora</i>	Categórica	Intervalo de horas de la posición de la gaviota
<i>fechaHoraO</i>	Fecha	Día inicio de la trayectoria
<i>fechaHoraU</i>	Fecha	Día final de la trayectoria
<i>DiaSem</i>	Categórica	Día de la semana que se inicia la trayectoria
<i>DistNido_m</i>	Numérica	Distancia al nido en metros
<i>LAT</i>	Numérica	Coordenada latitud de la posición
<i>LONG</i>	Numérica	Coordenada longitud de la posición
<i>velKmH</i>	Numérica	Velocidad de la gaviota en km/h
<i>habitat</i>	Binaria	Indica si la gaviota está en ricefield (0 o 1)
<i>Habitat2</i>	Binaria	Indica si la gaviota está en tierra (0 o 1)
<i>Colony</i>	Binaria	Indica si la gaviota está en la colonia (0 o 1)

2.2 Preproceso de los datos

El paso previo al análisis de trayectorias fue la fase de preproceso de los datos para no incluir ningún sesgo en el análisis posterior debido a una mala depuración de los datos.

Primero de todo se comprobó si realmente los tiempos entre dos localizaciones consecutivas era siempre de aproximadamente 5 minutos. Para ello se creó la variable *time_spent* a partir de la fecha de cada localización menos la fecha del punto anterior de la trayectoria para así saber el tiempo real en segundos que tarda el GPS en actualizar a un nuevo punto de la trayectoria.

El grafico de densidad de esta variable es el siguiente:

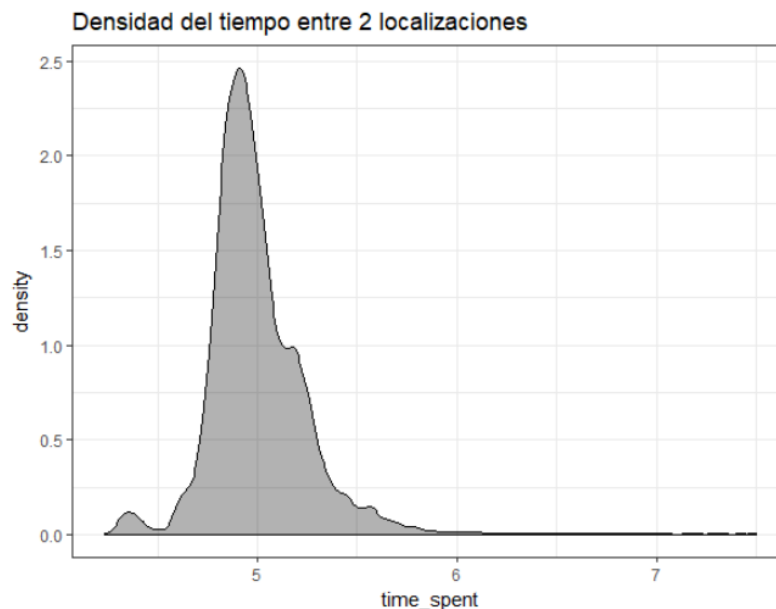


Figura 2.2: Distribución de la variable timespent

La mediana del tiempo es de 4.9 y la media de 5 minutos. En cuanto a la dispersión, la desviación obtenida es de 0.52 minutos. Por lo que no se puede asegurar que el GPS siempre se actualice exactamente cada 5 minutos, es decir, las trayectorias a considerar no tienen una longitud de paso totalmente constante. Esto no es ningún problema para el análisis de trayectorias, pero se debe tener en cuenta a la hora de construir los índices.

También se corrigieron los errores tipográficos de algunas categorías (error tipográfico en una de las categorías de *idViaje1* categoría incorrecta “V1b” que se modifica por “V01”) y los errores de formato en el caso de las variables *fechaOfici*, *horaOficia* se convierten en formato fecha y hora respectivamente.

No hizo falta una imputación de missings en los datos ya que la información para todas las variables estaba completa.

Además, se crearon las siguientes nuevas variables:

- **día:** número de días desde inicio del seguimiento
- **tiempos:** tiempo total acumulado desde el inicio de cada trayectoria.

Y a nivel trayectoria, estas otras:

- **tipo_día:** Indica si el día de salida es laborable o fin de semana.
- **duración:** Tiempo total de vuelo en minutos. Dado que hay trayectorias que pueden implicar más de 1 día, hizo falta crear otra variable que es la fecha completa (información del día y hora de cada localización)
- **momento:** Indica el momento del día de inicio de la trayectoria: mañana (de 6 a 13h), tarde (de las 14h a las 19h), noche (de las 20h a las 5h).

Finalmente, las variables que se excluyen del análisis, dado que no aportan ninguna información importante para las trayectorias son: *Campanya* (es variable constante), *intervHora*, *habitat*, *habitat2* y *Colony* (no son útiles para el análisis de caracterización de trayectorias).

2.3 Descriptiva

Número de trayectorias por gaviota

Se observa que el número de trayectorias no es el mismo para cada gaviota. Hay 2 gaviotas con solo 2 trayectorias y la media de número de trayectorias por gaviota es de 10.

En total hay 362 trayectorias: 297 en los días laborables (82% del total) y 65 en el fin de semana (18%).

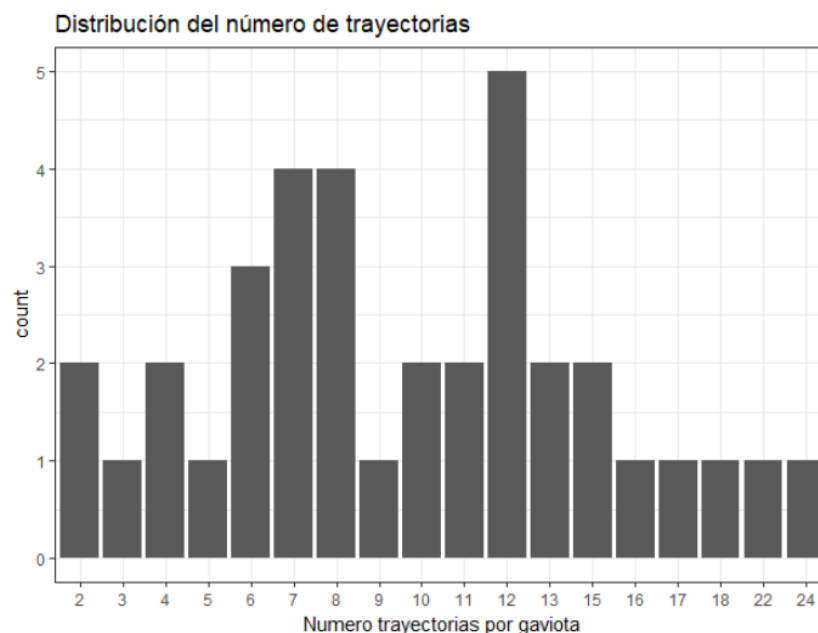


Figura 2.3: histograma del número de trayectorias por gaviota

Distancia máxima al nido de cada gaviota según día de la semana

Las trayectorias de las gaviotas en mediana no llegan a más de 50 km de lejanía respecto su nido (aunque como se ve en el gráfico hay algunas trayectorias que superan los 100km de distancia).

El viernes es el día con una distancia mediana máxima al nido y después estas distancias van disminuyendo en el fin de semana.

Cabe destacar que las distancias máximas de las trayectorias en el domingo son bastante menores en relación con otros días (el percentil 75 en el domingo es menor incluso que las distancias medianas de los otros días). Las conclusiones del estudio de composición del hábitat “*Spatiotemporal analyses on tracking data reveal fine-scale, daily cycles in seabird fisheries interactions*” dan una posible respuesta a esta gran diferencia: pasan mayor tiempo en el mar durante los días laborables debido a la actividad pesquera. En los festivos, se pasan menos tiempo en el mar, y van a los arrozales más cercanos.

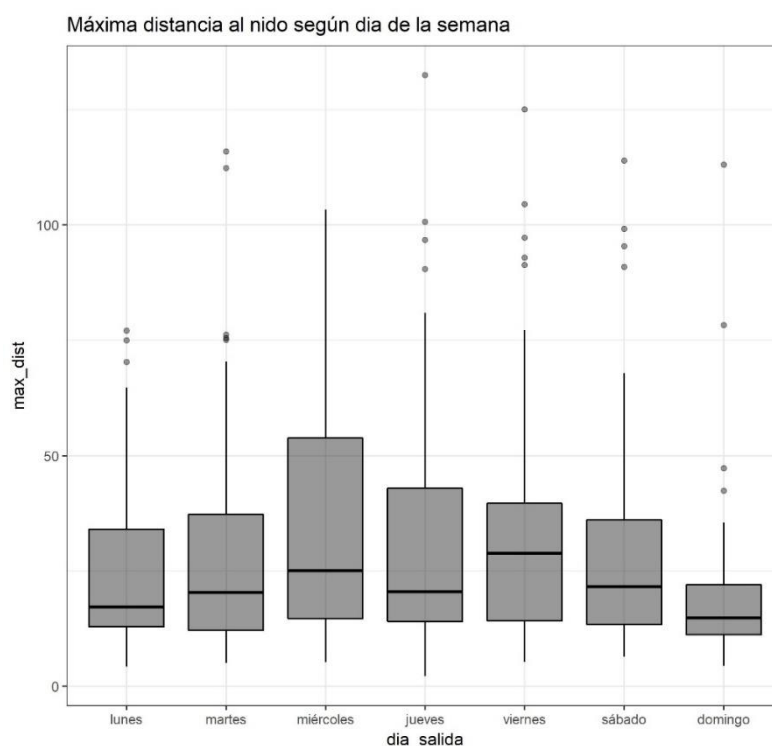


Figura 2.4: Distancia máxima al nido según día

Momento inicial de la trayectoria según el tipo de día

Destaca el sábado que es cuando el 60% de las trayectorias inician el vuelo por la mañana mientras que los otros días, las trayectorias de mañana no representan ni el 50%.

En los otros días, la proporción de trayectorias está más repartida. No es sorprendente ver que por la noche también vuelan ya que es de las pocas especies de gaviota con hábitos nocturnos desarrollados (Fuente: SEO Bird Life)

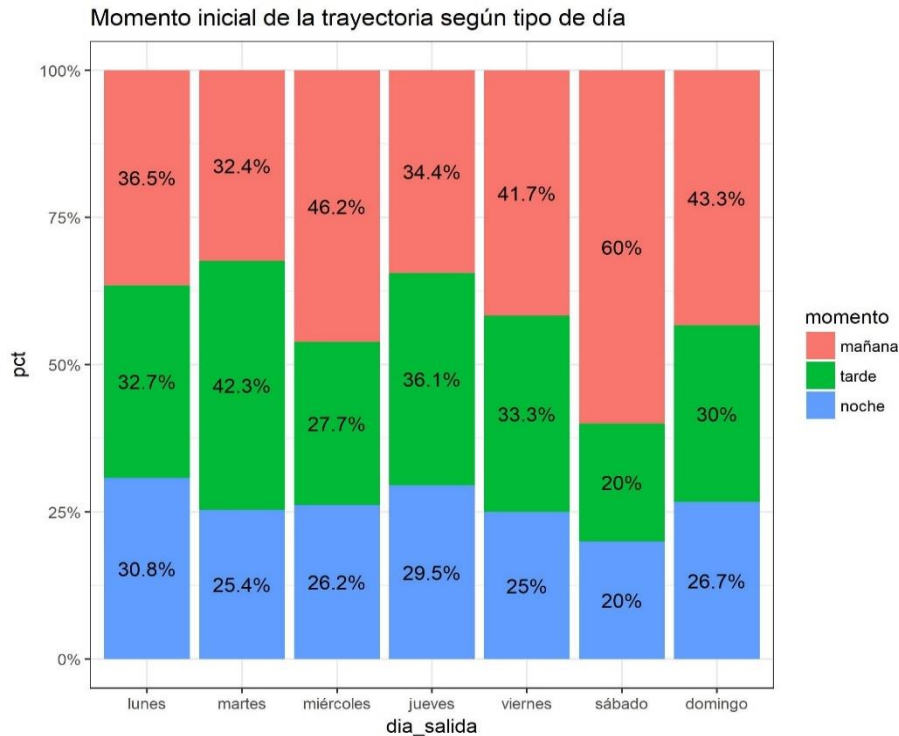


Figura 2.5: Momento inicial del vuelo según día

2.4 Obtención de los datos a nivel trayectoria

La función `TrajrFromCoords` del paquete `trajr` crea el objeto `trayectoria`. Se necesita para ello la información de las posiciones de cada punto (longitud y latitud) y la variable *tiempos* (el tiempo acumulado en cada punto de la trayectoria desde el inicio) Con este objeto, ya es posible crear los índices para la caracterización de las trayectorias.

La dimensión final del dataframe es de **362 filas** (trayectorias) y **10 columnas** (8 índices y la información del tipo y momento del día). Los índices finalmente calculados son: **índice de sinuosidad corregida, media de los cambios direccionales, desviación de los cambios direccionales, velocidad media, aceleración media, desviación de las velocidades, desviación de las aceleraciones y duración.**

Todos estos índices, a excepción de la duración se crean a partir de funciones implementadas en el paquete `trajr`. Para el cálculo de la sinuosidad se rediscretizó las trayectorias y para calcular las velocidades y aceleraciones se aplicó el filtro de Savitsky Golay (orden polinomio de 3)

Además, se adjunta la información del tipo de día de salida y el momento inicial del día para la modelización de cada índice en el siguiente capítulo.

CAPÍTULO 3: Modelos mixtos para el análisis de los índices

En esta parte del trabajo se analiza cada índice por separado mediante modelos lineales mixtos, dada la no independencia de los datos. Los gráficos de la validación se encuentran en el anexo.

3.1 Análisis de la sinuosidad

A continuación, se adjunta la distribución de la sinuosidad y la distribución después de aplicar la transformación logarítmica.

Finalmente, se decide trabajar con el logaritmo de la sinuosidad para que tenga una distribución más normal.

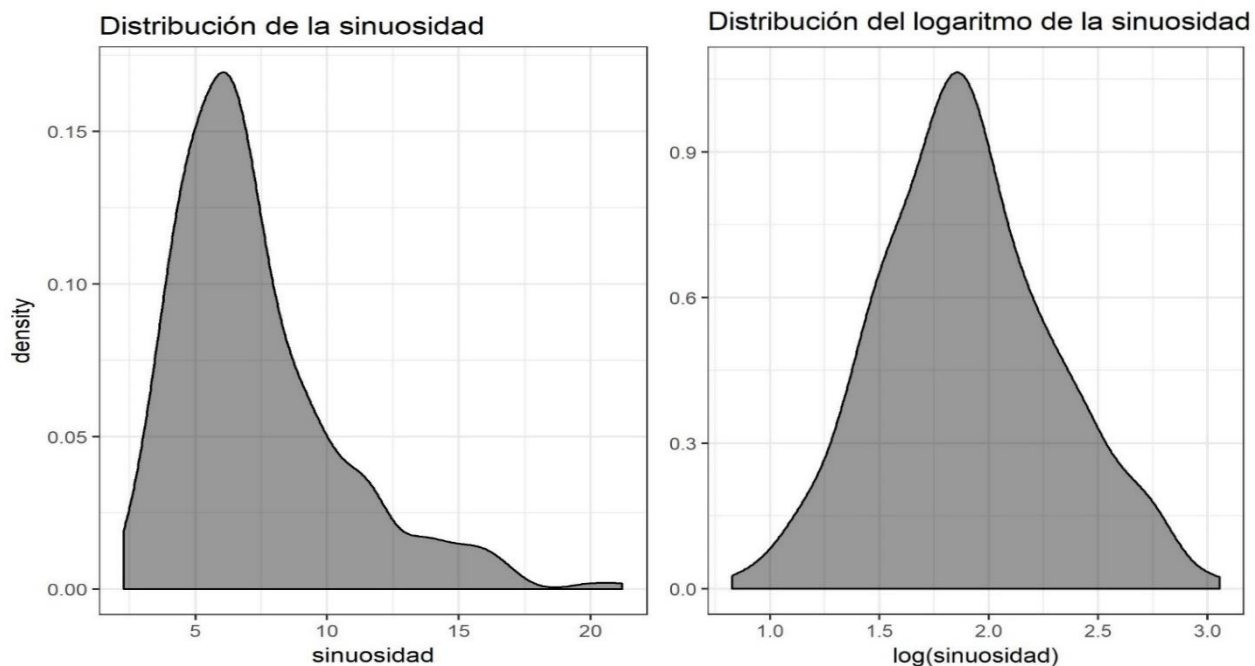


Figura 3.1 Distribución de sinuosidad y log sinuosidad

La media de sinuosidad es de 7.2 y la mediana de 6.5.

3.1.1 Determinación de la estructura de varianzas y covarianzas

Los resultados para determinar la estructura de covarianzas se muestran a continuación:

Modelo	Efecto aleatorio	Error residual	AIC
1	Homoscedástico	Homoscedástico	313.6
2	Homoscedástico	Heteroscedástico por tipo de día	315.5
3	Varianza entre gaviotas diferente según tipo de día	Homoscedástico	310
4	Varianza entre gaviotas diferente según tipo de día	Heteroscedástico por tipo de día	308.5

Tabla 3.1: Resultados AIC para determinar la estructura de varianzas y covarianzas

Los resultados del Likelihood Ratio Test son los siguientes:

Si comparamos el modelo 1 con el modelo 2, según el LRT test es mejor el modelo 1. Pero el modelo 3 es preferible al modelo 1 y también al modelo 4. Por lo que finalmente, el **mejor modelo es el que tiene varianza entre gaviotas diferente y varianza residual también diferente según el tipo de día**. Ambos criterios (AIC y LRT) coinciden en la elección del mejor modelo.

```
> anova(m1,m2)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 313.6102 325.2851 -153.8051
m2     2  4 315.4842 331.0507 -153.7421 1 vs 2 0.1260477 0.7226
> anova(m1,m3)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 313.6102 325.2851 -153.8051
m3     2  5 310.0220 329.4802 -150.0110 1 vs 2 7.588224 0.0225
> anova(m3,m4)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m3     1  5 310.0220 329.4802 -150.0110
m4     2  6 308.5275 331.8774 -148.2638 1 vs 2 3.494462 0.0616
```

Salida R 3.1: Resultados LRT test para determinar la estructura de varianzas y covarianzas

3.1.2 Introducción de efectos fijos

Los resultados, con la estructura elegida en el apartado anterior son los siguientes:

Modelo	Efectos fijos	AIC
5	Tipo de día	298
6	Tipo de día y momento del día	294.1
7	Interacción de los 2 efectos fijos	293.8

Tabla 3.2: Resultados AIC de los efectos fijos

El LRT test concluye que el modelo que incorpora los dos efectos fijos es mejor que el modelo 6 como se ve a continuación en el output. Pero el hecho de incorporar además la interacción entre estos dos efectos no mejora el ajuste del modelo según el LRT test (aunque según el AIC, sí).

Finalmente, se elige el modelo 6 como **mejor modelo** para explicar la sinuosidad de las trayectorias (**tipo de día y momento del día entran al modelo, pero sin estar interaccionados**)

```
> anova(m5,m6)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m5      1  7 298.0687 325.3103 -142.0344
m6      2  9 294.1208 329.1456 -138.0604 1 vs 2  7.947924  0.0188
> anova(m6,m7)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m6      1  9 294.1208 329.1456 -138.0604
m7      2 11 293.8080 336.6161 -135.9040 1 vs 2  4.312817  0.1157
```

Salida R 3.2: Resultados LRT test de los efectos fijos

El summary del mejor modelo es el siguiente:

```

> summary(m6)
Linear mixed-effects model fit by REML
Data: idxs
      AIC      BIC    logLik
311.694 346.6188 -146.847

Random effects:
Formula: ~1 + tipo_dia | id
Structure: General positive-definite, Log-Cholesky parametrization
              StdDev   Corr
(Intercept)  0.2888417 (Intr)
tipo_dialaborable 0.1637219 -0.66
Residual      0.2764500

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | tipo_dia
Parameter estimates:
fin_de_semana    laborable
1.000000         1.204668
Fixed effects: log(sinuosidad) ~ tipo_dia + momento
              Value Std.Error DF   t-value p-value
(Intercept)  2.0147367 0.06493675 323 31.026140 0.0000
tipo_dialaborable -0.2134898 0.05159972 323 -4.137421 0.0000
momentotarde      0.1206526 0.04378625 323  2.755490 0.0062
momentonoche      0.0214292 0.04687596 323  0.457146 0.6479
Correlation:
              (Intr) tp_dlb mmnttr
tipo_dialaborable -0.676
momentotarde      -0.239 -0.100
momentonoche      -0.239 -0.083  0.440

Standardized Within-Group Residuals:
              Min              Q1              Med              Q3              Max
-2.76743424 -0.59773031  0.01796987  0.65676263  2.84336235

Number of Observations: 362
Number of Groups: 36

```

Salida R 3.3: Summary del mejor modelo

```

> anova(m6)
              numDF denDF   F-value p-value
(Intercept)      1   323 1990.7417 <.0001
tipo_dia          1   323  15.3863 0.0001
momento           2   323   4.1506 0.0166

```

Salida R 3.4: Significación de los efectos fijos

En la función de varianza, se especifica que en días laborables hay un 20% más de variabilidad residual respecto los días festivos.

El hecho de que una **gaviota vuele en día laborable o fin de semana y que empiece a volar en diferentes momentos del día** son dos factores estadísticamente significativos a la hora de explicar la sinuosidad de la trayectoria (ambos coeficientes presentan p-valores < 0.05).

Hay un decrecimiento, en media, del logaritmo de la sinuosidad de -0.21 (IC 95%: -0.31, -0.11) en los días laborables respecto al fin de semana. En escala proporcional: la sinuosidad en los días laborables es un 23.3% inferior respecto el fin de semana. La desviación del efecto aleatorio es de 0.16 que indica una gran variabilidad de este efecto entre las gaviotas.

En cuanto al momento del día, apenas hay cambio en el índice de una trayectoria de noche respecto a una de día (coeficiente momento noche=0.02 e IC 95%: -0.07, 0.11) y no es significativa. En escala proporcional, la sinuosidad es 2.02% mayor respecto el momento del día.

El cambio significativo es en la tarde: El cambio es mayor de día a tarde donde el coeficiente aumenta en 0.12 (IC 95%: 0.03, 0.20). En escala proporcional, la sinuosidad por la tarde es un 12.7% mayor respecto el día.

En cuanto al diagnóstico, se observa que los residuos obtenidos con el mejor modelo se comportan como una distribución normal de media 0. Además, ninguna observación presenta un residuo mayor a 3 en valor absoluto por lo que parece indicar que no existen atípicos.

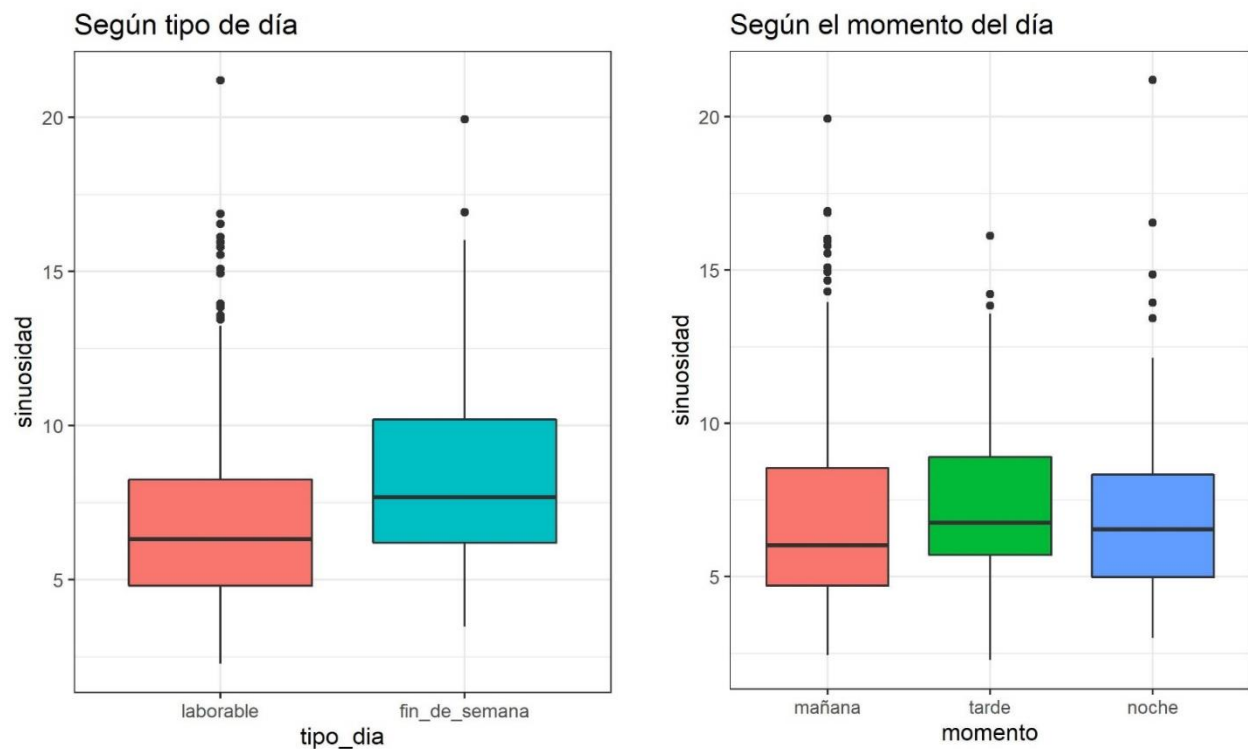


Figura 3.2: Sinuosidad según tipo de día y según momento del día

3.1.3 Repetibilidad del índice de sinuosidad

Para el índice de sinuosidad, el ICC vale 0.29 (IC 95%: 0.14, 0.41) lo que indica una repetibilidad baja pero distinta de 0.

Si desglosamos este ICC según el tipo de día tenemos que el ICC de sinuosidad para los días festivos vale 0.52 (IC 95%: 0.21, 0.73) y para los días laborables vale 0.29 (IC 95%: 0.14, 0.43). Esto quiere decir que, en los días festivos, los valores de sinuosidad presentan mayor fiabilidad en comparación con los de días laborables.

Pero si se realiza el test de permutaciones, el p-valor obtenido con 1000 permutaciones es de 0.107 por lo que no se rechaza la igualdad de ICC, es decir, en realidad la diferencia no es estadísticamente significativa en cuanto a la consistencia de la sinuosidad según el tipo de día.

```
> sum(abs(f.perms) > abs(dif_original))/nperm # two-tailed test  
[1] 0.1071429
```

Salida R 3.5: Resultado test de permutaciones

3.2 Análisis de la duración

Al igual que con la sinuosidad, se aplica la transformación logarítmica:

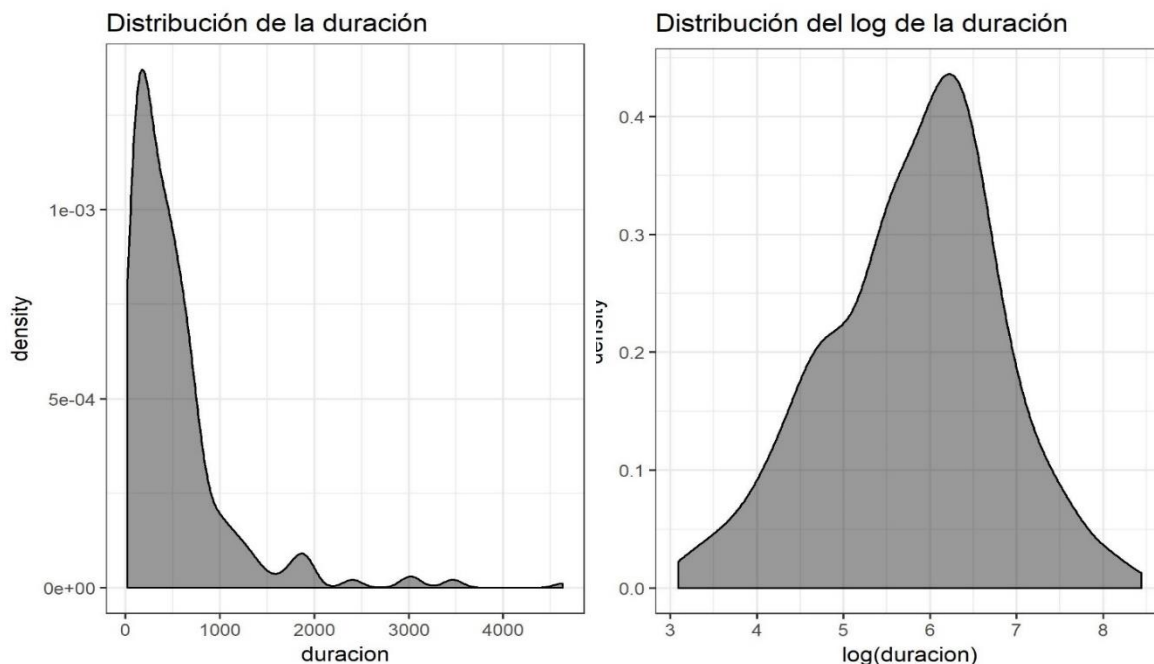


Figura 3.3: Distribución de la duración y log duración

Las gaviotas vuelan en mediana 6.3 horas. La media se ve afectada por los valores altos (hay una trayectoria que el vuelo tiene 77.1 horas de duración. Corresponde a la trayectoria 510794V04 que se adjunta a continuación:

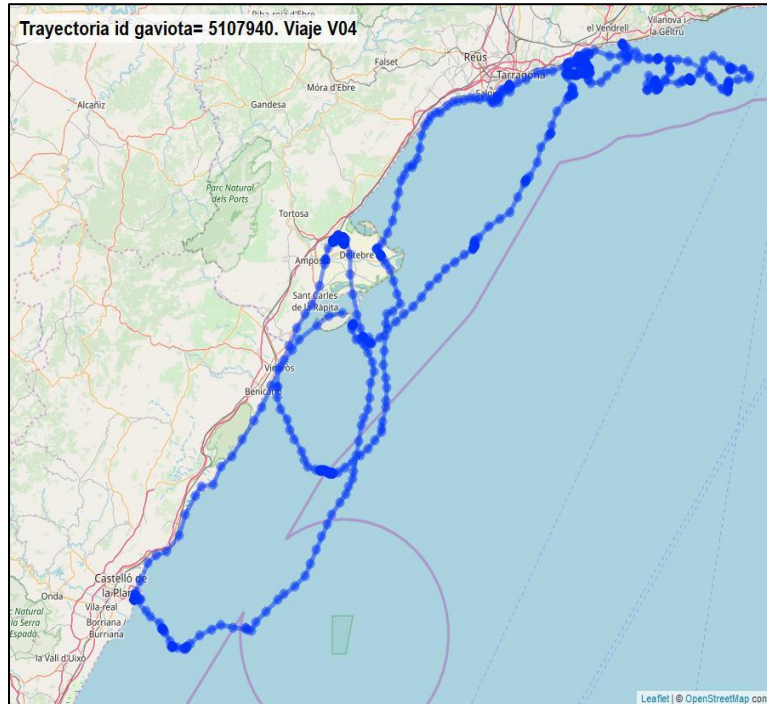


Figura 3.4: Visualización de la trayectoria 5107940V04

3.2.1 Determinación de la estructura de varianzas y covarianzas

Modelo	Efecto aleatorio	Error residual	AIC
1	Homoscedástico	Homoscedástico	1015.5
2	Homoscedástico	Heteroscedástico por tipo de día	1011.1
3	Varianza entre gaviotas diferente según tipo de día	Homoscedástico	1014.1
4	Varianza entre gaviotas diferente según tipo de día	Heteroscedástico por tipo de día	1014.7

Tabla 3.3: Resultados AIC para determinar la estructura de varianzas y covarianzas

La mejor estructura de varianzas y covarianzas es especificar una **varianza entre gaviotas igual, pero varianza residual diferente según tipo de día**. Ambos criterios llevan a la misma conclusión.

```

> anova(m1,m2)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 1015.530 1027.205 -504.7652
m2     2  4 1011.087 1026.653 -501.5432 1 vs 2 6.44403 0.0111
> anova(m2,m3)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m2     1  4 1011.087 1026.653 -501.5432
m3     2  5 1014.137 1033.595 -502.0686 1 vs 2 1.050649 0.3054
> anova(m2,m4)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m2     1  4 1011.087 1026.653 -501.5432
m4     2  6 1014.753 1038.103 -501.3765 1 vs 2 0.333355 0.8465

```

Salida R 3.6: Resultados LRT test para determinar la estructura de varianzas y covarianzas

3.2.2 Introducción de efectos fijos

Se introducen los efectos fijos al modelo con la mejor estructura. Los resultados son los siguientes:

Modelo	Efectos fijos	AIC
5	Tipo de día	1006.3
6	Tipo de día y momento del día	991.6
7	Interacción de los 2 efectos fijos	994.3

Tabla 3.4: Resultados AIC de los efectos fijos

```

> anova(m5,m6)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m5     1  5 1006.3280 1025.786 -498.1640
m6     2  7  991.5811 1018.823 -488.7906 1 vs 2 18.74684 1e-04
> anova(m6,m7)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m6     1  7  991.5811 1018.823 -488.7906
m7     2  9  994.2806 1029.305 -488.1403 1 vs 2 1.300539 0.5219

```

Salida R 3.7: Resultados LRT test de los efectos fijos

Ambos criterios concluyen que el **mejor modelo es el que incluye ambas variables sin interacción.**

A continuación, el summary del mejor modelo:


```

> summary(m6)
Linear mixed-effects model fit by maximum likelihood
Data: idxs[-171, ]
      AIC      BIC    logLik
892.218 1009.44 -484.109

Random effects:
Formula: ~1 | id
      (Intercept) Residual
StdDev:   0.3171775  1.04839

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | tipo_dia
Parameter estimates:
fin_de_semana    laborable
      1.0000000    0.8182528
Fixed effects: log(duracion) ~ tipo_dia + momento
              Value Std.Error DF t-value p-value
(Intercept)    6.383479 0.1516523 322 42.09286 0.0000
tipo_dialaborable -0.326795 0.1418630 322 -2.30360 0.0219
momentotarde     -0.525684 0.1162000 322 -4.52396 0.0000
momentonoche     -0.339882 0.1238876 322 -2.74347 0.0064
Correlation:
              (Intr) tp_dlb mmnttr
tipo_dialaborable -0.779
momentotarde      -0.276 -0.090
momentonoche      -0.271 -0.075  0.442

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3      Max
-3.27560056 -0.65389863  0.08227924  0.68868776  2.43387843

Number of Observations: 361
Number of Groups: 36

```

Salida R 3.8: Summary del mejor modelo

```

> anova(m6)
              numDF denDF F-value p-value
(Intercept)      1   323 5769.84 <.0001
tipo_dia          1   323   7.52 0.0064
momento           2   323   9.54 0.0001

```

Salida R 3.9: Significación de los efectos fijos

En el apartado de función de varianza, destaca los días laborables con una variabilidad total residual un 22% inferior (1/0.818) respecto a la que hay en los días festivos.

Si nos centramos en la significación de los efectos, se observa que ambos efectos fijos son estadísticamente significativos para explicar la duración de las trayectorias. El log de la duración disminuye en 0.32 (IC 95%: -0.6, -0.04) si se trata de una trayectoria en día laborable respecto a una que sucede en día festivo (en escala proporcional; la media de duración del vuelo es un 38% inferior a la de los festivos). Pero la pendiente es aleatoria con una desviación importante entre gaviotas de 0.31.

Según cada momento del día, la tarde es donde hay un decremento mayor respecto al día (coeficiente momento tarde de -0.52. IC 95%: -0.75, -0.3) y en el caso de la noche el decremento es de -0.33 (IC 95%: -0.58, -0.09). En escala proporcional, la duración es un 63% inferior a la tarde y un 39% inferior a la noche respecto las trayectorias de día.

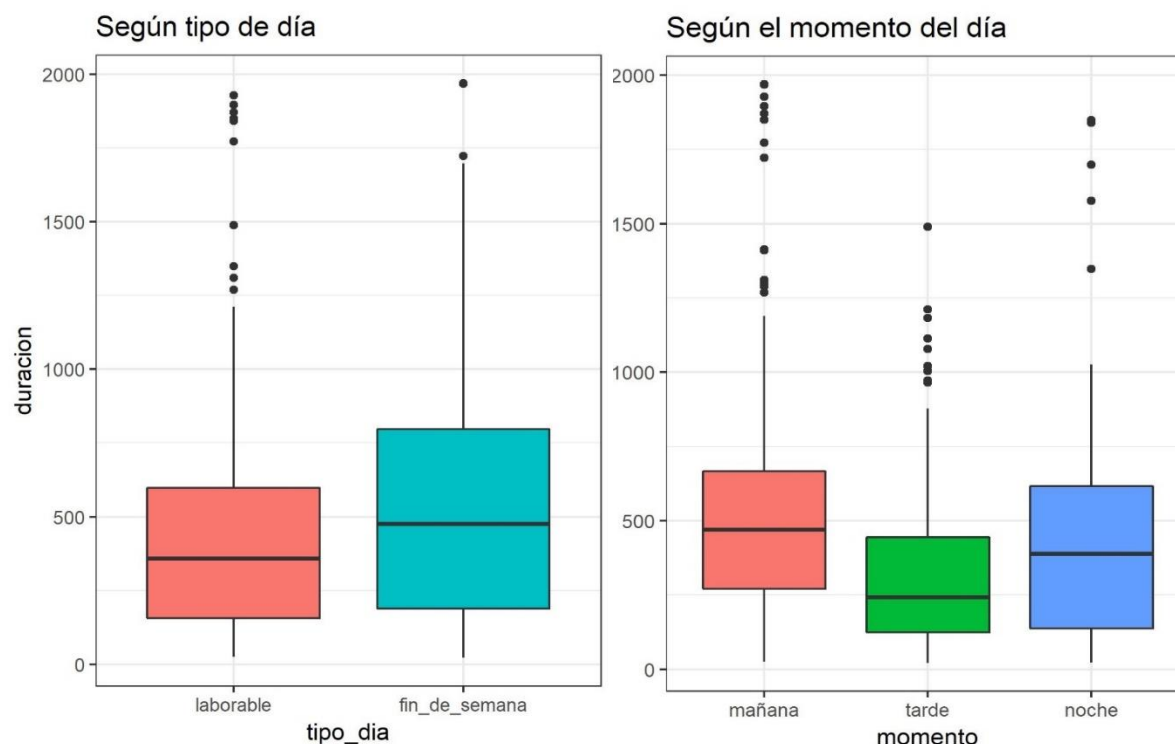


Figura 3.5: Duración según tipo de día y según momento del día

En cuanto a la validación, se observa que los residuos obtenidos con el mejor modelo se comportan como una distribución normal de media 0. Además, ninguna observación presenta un residuo mayor a 3 en valor absoluto, a excepción únicamente de una trayectoria que presenta un residuo de -3.2 correspondiente a un vuelo que dura solo 24 minutos. Se reestima el modelo sin esta observación (el summary de este modelo se encuentra en el anexo):

Se observa que el AIC pasa de 982 a 978, los coeficientes cambian un poco, pero mantienen las mismas relaciones que en el primer modelo y ahora ya todos los residuos son menores a 3 en valor absoluto.

3.2.3 Repetibilidad de la duración

El ICC de duración vale 0.12 (IC 95%: 0.02, 0.21) lo que indica una **repetibilidad relativamente baja**.

Al ser un ICC muy cercano a 0.1 se realiza un test para ver si en realidad es estadísticamente igual a 0 o no comparando el modelo con efecto aleatorio con uno sin efecto aleatorio (modelo lineal no mixto). A pesar de tener una repetibilidad muy baja, esta es **distinta estadísticamente**

de 0 (p-valor<0.05). Por lo tanto, hay comportamiento diferente entre las gaviotas en cuanto a la duración de la trayectoria.

```
> anova(m_nomixt,m_mixt)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m_nomixt   1  2 1030.392 1038.175 -513.1959
m_mixt     2  3 1015.530 1027.205 -504.7652 1 vs 2 16.86132 <.0001
```

Salida R 3.10: Resultado test de significación ICC

El ICC de la duración es de 0.10 (IC 95%: 0.01, 0.19) para los días laborables y de 0.25 (IC 95%: 0, 0.6) para días festivos. Si se aplica el test de permutaciones, se concluye que **no existen diferencias significativas según el tipo de día** (p-valor=0.21 >0.05).

```
> sum(abs(f.perms) > abs(dif_original))/nperm # two-tailed test
[1] 0.214
```

Salida R 3.11: Resultado test de permutaciones

3.3 Análisis de la no linealidad (media del cambio direccional)

En este caso, no se aplica la transformación logarítmica. La distribución de la media del cambio direccional es la siguiente:

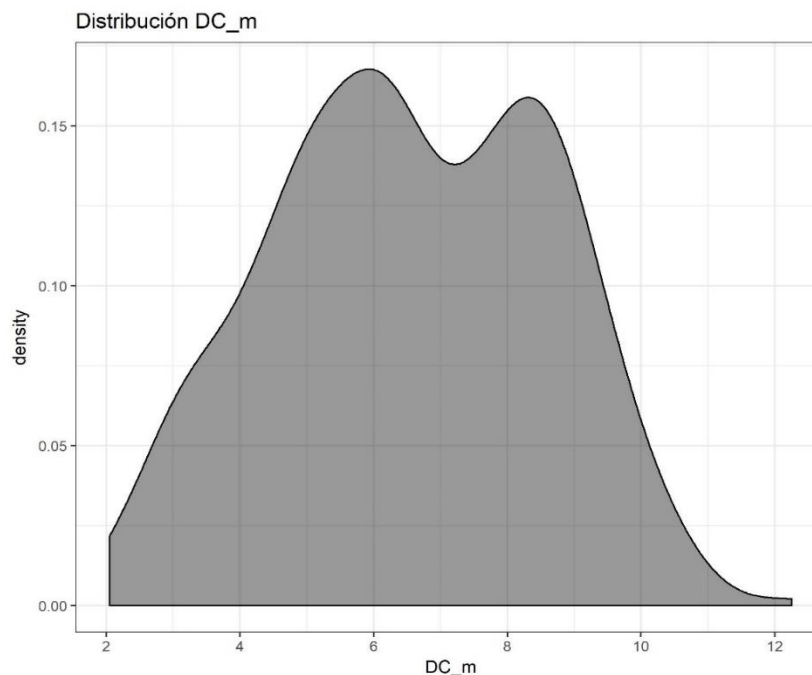


Figura 3.6: Distribución media del cambio direccional

3.3.1 Determinación de la estructura de varianzas y covarianzas

Modelo	Efecto aleatorio	Error residual	AIC
1	Homoscedástico	Homoscedástico	1518.4
2	Homoscedástico	Heteroscedástico por tipo de día	1518.2
3	Varianza entre gaviotas diferente según tipo de día	Homoscedástico	1518.7
4	Varianza entre gaviotas diferente según tipo de día	Heteroscedástico por tipo de día	1514.3

Tabla 3.5: Resultados AIC para determinar la estructura de varianzas y covarianzas

```

> anova(m1,m2)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 1518.376 1530.051 -756.1878
m2     2  4 1518.189 1533.756 -755.0946 1 vs 2 2.186329 0.1392
> anova(m1,m3)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 1518.376 1530.051 -756.1878
m3     2  5 1518.768 1538.226 -754.3841 1 vs 2 3.607318 0.1647
> anova(m1,m4)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 1518.376 1530.051 -756.1878
m4     2  6 1514.331 1537.680 -751.1653 1 vs 2 10.04504 0.0182

```

Salida R 3.12: Resultados LRT test para determinar la estructura de varianzas y covarianzas

Ambos criterios concluyen que la estructura óptima en cuanto al efecto aleatorio es una **varianza entre diferente según el tipo de día**, y una **varianza residual también diferente**.

3.3.2 Introducción de efectos fijos

Modelo	Efectos fijos	AIC
5	Tipo de día	1507.4
6	Tipo de día y momento del día	1494.1
7	Interacción de los 2 efectos fijos	1491

Tabla 3.6: Resultados AIC de los efectos fijos

```
> anova(m5,m6)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
m5	1	7	1507.413	1534.655	-746.7065			
m6	2	9	1494.086	1529.111	-738.0429	1 vs 2	17.3272	2e-04

```
> anova(m6,m7)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
m6	1	9	1494.086	1529.111	-738.0429			
m7	2	11	1491.058	1533.867	-734.5292	1 vs 2	7.027339	0.0298

Salida R 3.13: Resultados LRT test de los efectos fijos

El **mejor modelo** es el que incluye tanto el **tipo de día y el momento interaccionados**. Tanto el AIC como el LRT test concluyen lo mismo.

El summary del modelo es el siguiente:

```

> summary(m7)
Linear mixed-effects model fit by REML
Data: idxs
      AIC      BIC    logLik
1492.858 1535.482 -735.429

Random effects:
Formula: ~1 + tipo_dia | id
Structure: General positive-definite, Log-Cholesky parametrization
              StdDev   Corr
(Intercept)   0.8414843 (Intr)
tipo_dialaborable 0.6663657 -0.34
Residual      1.4280810

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | tipo_dia
Parameter estimates:
fin_de_semana      laborable
      1.000000      1.263348
Fixed effects: DC_m ~ tipo_dia * momento
              Value Std.Error  DF   t-value p-value
(Intercept)   7.377887 0.2993122 321  24.649468  0.0000
tipo_dialaborable -1.504184 0.3338460 321  -4.505621  0.0000
momentotarde    -0.228261 0.4746111 321  -0.480943  0.6309
momentonoche     0.031779 0.4910169 321   0.064721  0.9484
tipo_dialaborable:momentotarde 1.558227 0.5359158 321   2.907597  0.0039
tipo_dialaborable:momentonoche 0.190991 0.5589859 321   0.341673  0.7328
Correlation:
              (Intr) tp_dlb mmnttr mmntnc tp_dlbbrbl:mmntt
tipo_dialaborable -0.732
momentotarde      -0.495  0.439
momentonoche      -0.488  0.424  0.314
tipo_dialaborable:momentotarde 0.428 -0.561 -0.874 -0.268
tipo_dialaborable:momentonoche 0.422 -0.545 -0.274 -0.866  0.346

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-2.2982525 -0.7420680  0.0264061  0.6893502  3.2390927

Number of Observations: 362
Number of Groups: 36

```

Salida R 3.14: Summary del mejor modelo

```

> anova(m7)
              numDF denDF   F-value p-value
(Intercept)       1   321  1574.8569 <.0001
tipo_dia           1   321   14.9935  0.0001
momento            2   321   10.3553 <.0001
tipo_dia:momento   2   321    4.4774  0.0121

```

Salida R 3.15: Significación de los efectos fijos

Nos indica que la varianza residual en los días laborables es un 26.3% más alta respecto a la variabilidad residual de los días festivos.

Se observa que todos los términos del modelo son significativos a la hora de explicar la no linealidad de las trayectorias (en la salida R 3.15 todos los p-valores < 0.05)

Una trayectoria en día laborable tiene un decremento de -1.5 en la media del cambio direccional respecto a una en día festivo pero este decremento se anula prácticamente si la trayectoria además sucede por la tarde (coeficiente de la interacción laborable:tarde = 1.55. IC 95%: 0.5, 2.6) y es cuando no hay apenas diferencias según el tipo de día si sucede por la tarde.

Para trayectorias por la noche, la interacción entre día laborable y momento del día no logra anular el decremento, pero reduce un poco (coeficiente laborable_noche = 0.19. IC 95%: -0.9, 1.3) el gap según tipo de día.

Gráficamente, se observa las diferencias, que el modelo logra captar, según cada factor y la interacción.

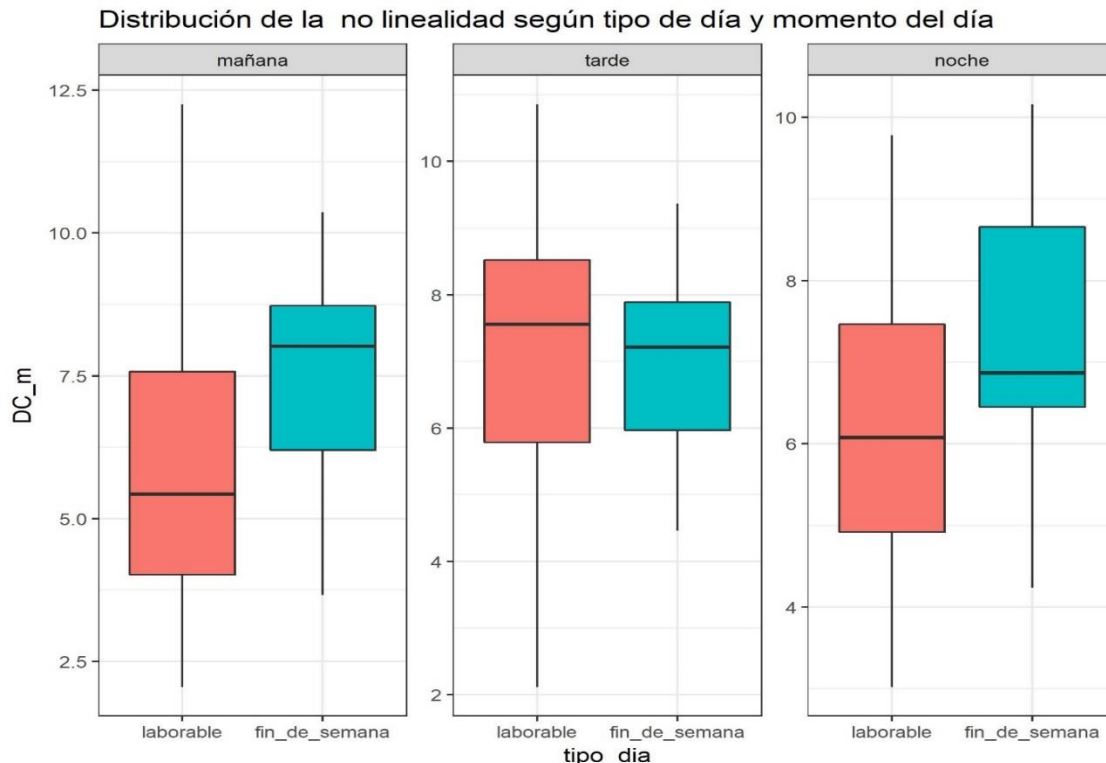


Figura 3.7: No linealidad según tipo y momento del día

En cuanto a la validación: errores normales centrados en 0, y los residuos estandarizados inferiores a 3 en valor absoluto (a excepción de un solo valor que tiene un residuo estandarizado de 3.23).

3.3.3 Repetibilidad de la no linealidad

El ICC del índice de no linealidad vale **0.16** (IC 95%: 0.05, 0.26) lo que indica una **repetibilidad relativamente baja**. A pesar de esto, es **estadísticamente diferente de 0** (p-valor LRT test <0.05). Por lo tanto, hay comportamiento diferente entre las gaviotas en cuanto a la duración de la trayectoria.

```
> anova(m_nomixt,m_mixt)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m_nomixt   1  2 1550.319 1558.102 -773.1594
m_mixt     2  3 1523.129 1534.804 -758.5645 1 vs 2 29.18977 <.0001
```

Salida R 3.16: Resultado test de significación ICC

El ICC de la no linealidad es de 0.18 (IC 95%: 0.06, 0.29) para los días laborables y de 0.29 (IC 95%: 0, 0.55) para días festivo. Si se aplica el test de permutaciones, se concluye que **no existen diferencias significativas según el tipo de día** (p-valor=0.45>0.05).

```
> sum(abs(f.perms) > abs(dif_original))/nperm # two-tailed test
[1] 0.454
```

Salida R 3.17: Resultado test de permutaciones

3.4 Análisis de la irregularidad (desviación típica del cambio direccional)

Si se analiza la distribución de este índice se detecta la presencia de algunos valores bastantes elevados y si se aplica el logaritmo, el valor extremo sigue estando muy a la cola. Se trata de la trayectoria 5107959V06 que presenta un valor de DC_sd igual a 31.

Finalmente se decide eliminar esta trayectoria para el análisis de este índice para así no empeorar el ajuste de los modelos por presencia de atípicos.

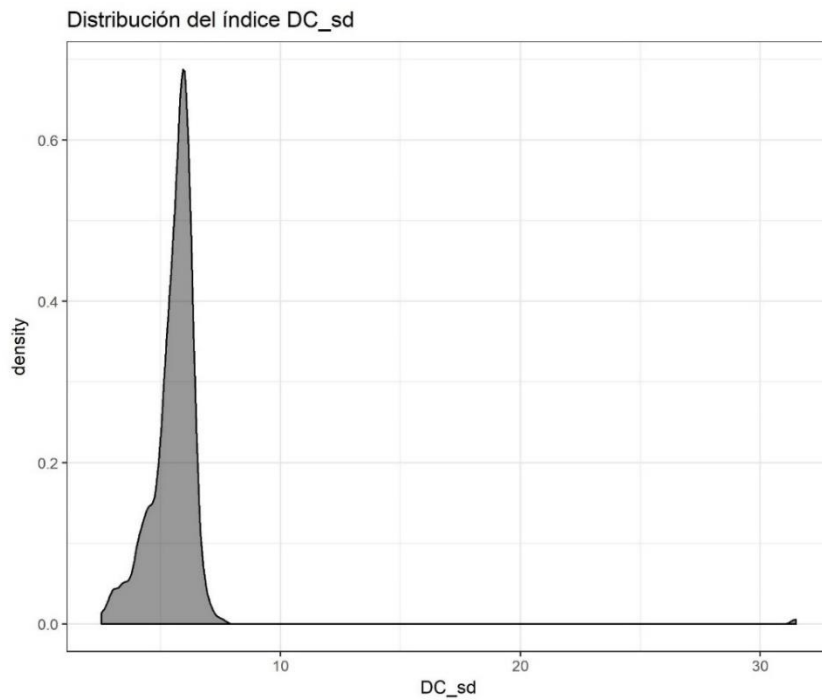


Figura 3.8: Distribución del índice DC_sd

La trayectoria con el valor atípico de DC_sd es la siguiente:

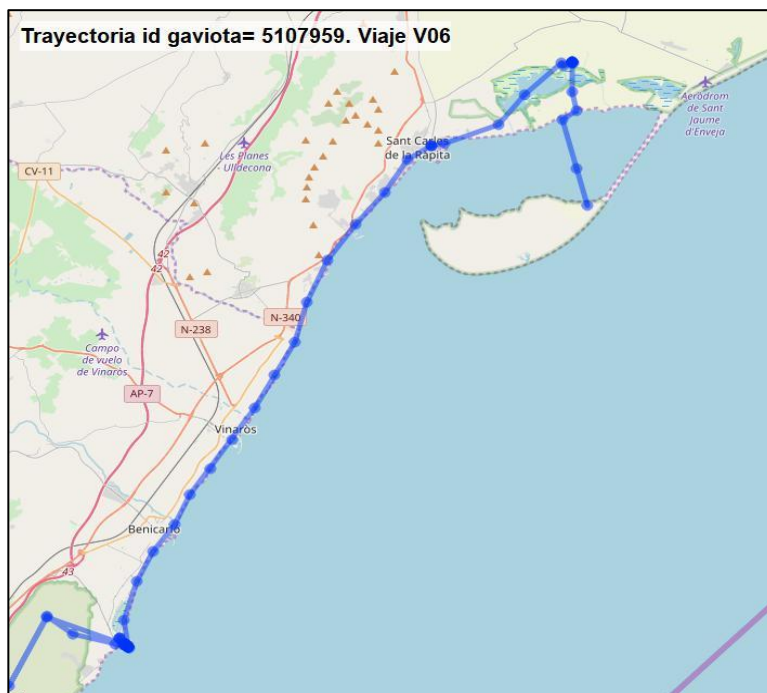


Figura 3.9: Visualización de la trayectoria 5107959V06

Una vez eliminada este dato tan atípico, la distribución queda de la siguiente manera.

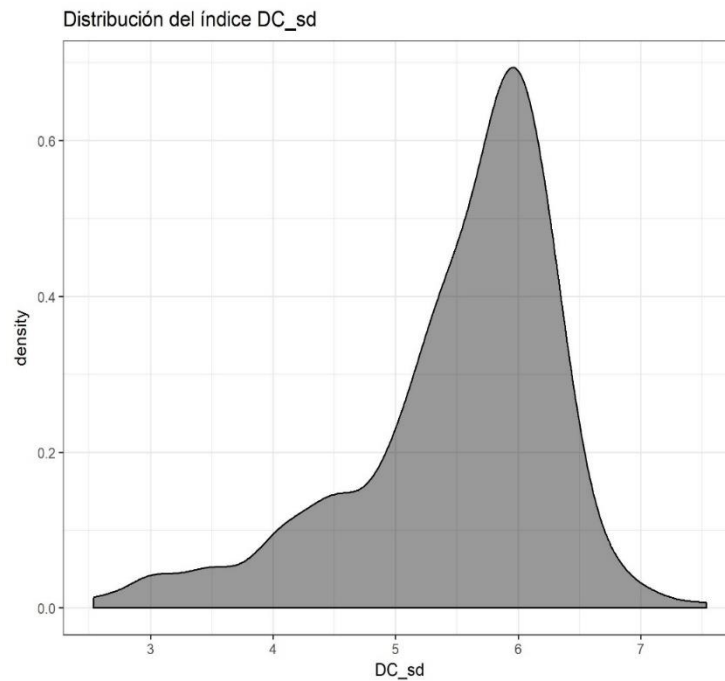


Figura 3.10: Distribución del índice DC sd sin el dato atípico

3.4.1 Determinación de la estructura de varianzas y covarianzas

Modelo	Efecto aleatorio	Error residual	AIC
1	Homoscedástico	Homoscedástico	887.8
2	Homoscedástico	Heteroscedástico por tipo de día	881.8
3	Varianza entre gaviotas diferente según tipo de día	Homoscedástico	889.1
4	Varianza entre gaviotas diferente según tipo de día	Heteroscedástico por tipo de día	880.9

Tabla 3.7: Resultados AIC para determinar la estructura de varianzas y covarianzas

La mejor estructura, al igual que con el índice de la media de cambio direccional es una **varianza residual diferente según tipo de día y varianza entre gaviotas también diferente**. Ambos criterios llevan a esta misma conclusión.

```

> anova(m1,m2)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 887.7895 899.4562 -440.8948
m2     2  4 881.8389 897.3944 -436.9195 1 vs 2  7.950606  0.0048
> anova(m2,m3)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m2     1  4 881.8389 897.3944 -436.9195
m3     2  5 889.0945 908.5389 -439.5472 1 vs 2  5.255566  0.0219
> anova(m3,m4)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m3     1  5 889.0945 908.5389 -439.5472
m4     2  6 880.9254 904.2586 -434.4627 1 vs 2 10.16911  0.0014

```

Salida R 3.18: Resultados LRT test para determinar la estructura de varianzas y covarianzas

3.4.2 Introducción de efectos fijos

Se estiman 3 modelos con la mejor estructura de varianzas obtenida en el apartado anterior:

Modelo	Efectos fijos	AIC
5	Tipo de día	873.5
6	Tipo de día y momento del día	861.1
7	Interacción de los 2 efectos fijos	858

Tabla 3.8: Resultados AIC de los efectos fijos

```

> anova(m5,m6)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m5     1  7 873.5159 900.7380 -429.7580
m6     2  9 861.0976 896.0975 -421.5488 1 vs 2 16.4183  3e-04
> anova(m6,m7)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m6     1  9 861.0976 896.0975 -421.5488
m7     2 11 858.0239 900.8015 -418.0119 1 vs 2  7.073744  0.0291

```

Salida R 3.19: Resultados LRT test de los efectos fijos

El **mejor modelo** para explicar la irregularidad es el que incluye la **interacción de ambos factores**.

El summary de este modelo es el siguiente:

```

> summary(m7)
Linear mixed-effects model fit by REML
Data: idxs2
      AIC      BIC    logLik
873.3411 915.9344 -425.6705

Random effects:
Formula: ~1 + tipo_dia | id
Structure: General positive-definite, Log-Cholesky parametrization
              StdDev   Corr
(Intercept)  0.2661033 (Intr)
tipo_dialaborable 0.2388013 -0.426
Residual      0.5313397

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | tipo_dia
Parameter estimates:
fin_de_semana    laborable
1.000000         1.501402
Fixed effects: DC_sd ~ tipo_dia * momento
              Value Std.Error DF t-value p-value
(Intercept)  5.765425 0.1071871 320 53.78844 0.0000
tipo_dialaborable -0.576029 0.1300850 320 -4.42810 0.0000
momentotarde      0.193413 0.1751368 320 1.10435 0.2703
momentonoche      -0.206457 0.1808828 320 -1.14139 0.2546
tipo_dialaborable:momentotarde 0.306300 0.2068684 320 1.48065 0.1397
tipo_dialaborable:momentonoche 0.556438 0.2161296 320 2.57456 0.0105
Correlation:
              (Intr) tp_dlb mmnttr mmntnc tp_dlbrbl:mmntt
tipo_dialaborable -0.719
momentotarde      -0.511 0.416
momentonoche      -0.503 0.403 0.314
tipo_dialaborable:momentotarde 0.422 -0.575 -0.836 -0.256
tipo_dialaborable:momentonoche 0.414 -0.557 -0.261 -0.825 0.355

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-3.2382082 -0.4691605 0.1913899 0.6194175 2.0238419

Number of Observations: 361
Number of Groups: 36

```

Salida R 3.20: Summary del mejor modelo

```

> anova(m7)
              numDF denDF  F-value p-value
(Intercept)      1   320 9456.566 <.0001
tipo_dia          1   320  11.440 0.0008
momento           2   320   8.398 0.0003
tipo_dia:momento  2   320   3.578 0.0291

```

Salida R 3.21: Significación de los efectos fijos

Todos los coeficientes son estadísticamente significativos.

La irregularidad es menor en días laborables (coeficiente de -0.57 respecto a días festivos. IC 95%: -0.8, -0.3) pero este gap se reduce al completo si además la trayectoria en día laborable sucede por la noche ya que el coeficiente de la interacción es de 0.55 (IC 95%: 0.13, 0.98).

Por la tarde, en días laborable, el gap de 0.57 en relación con los días festivos, logra disminuir en media, pero no se anula (el coeficiente de esta interacción es de 0.3. IC 95%: -0.1, 0.7) y además la interacción por la tarde no es significativa.

En el apartado de función de varianza, se observa en el modelo que la variabilidad residual del índice en días laborables es un 50% mayor que en los días festivos.

Destaca también la gran variabilidad del índice que hay en las mañanas de los días laborables respecto las otras situaciones.

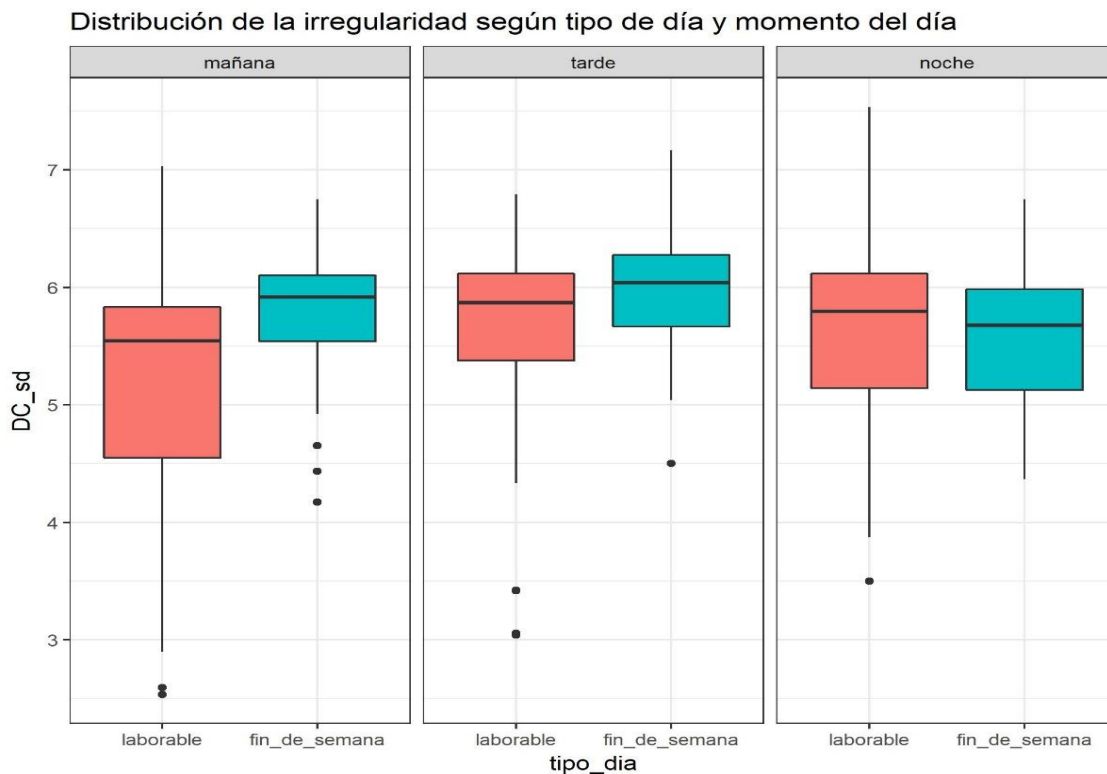


Figura 3.11: Irregularidad según tipo y momento del día

En cuanto a la validación del modelo, solo 2 trayectorias presentan residuos estandarizados superiores a 3 en valor absoluto.

3.4.3 Repetibilidad de la irregularidad

El **ICC para el índice de no linealidad vale 0.1** (IC 95%: 0.02, 0.18) lo que indica una repetibilidad relativamente baja. El test asegura que es **estadísticamente diferente de 0**. (p-valor<0.05)

```
> anova(m_nomixt,m_mixt)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
m_nomixt	1	2	1371.853	1379.636	-683.9263			
m_mixt	2	3	1367.231	1378.906	-680.6154	1 vs 2	6.621868	0.0101

Salida R 3.22: Resultado test de significación ICC

Por lo tanto, hay comportamiento diferente entre las gaviotas en cuanto a la irregularidad de la trayectoria. El ICC de la no linealidad es de 0.11 (IC 95%: 0.01, 0.2) para los días laborables y de 0.20 (IC 95%:0, 0.47) para días festivo. Sin embargo, si se aplica el test de permutaciones, se concluye que no existen diferencias significativas según el tipo de día (p-valor=0.43>0.05).

```
> sum(abs(f.perms) > abs(dif_original))/nperm # two-tailed test  
[1] 0.435
```

Salida R 3.23: Resultado test de permutaciones

3.5 Análisis de la velocidad media

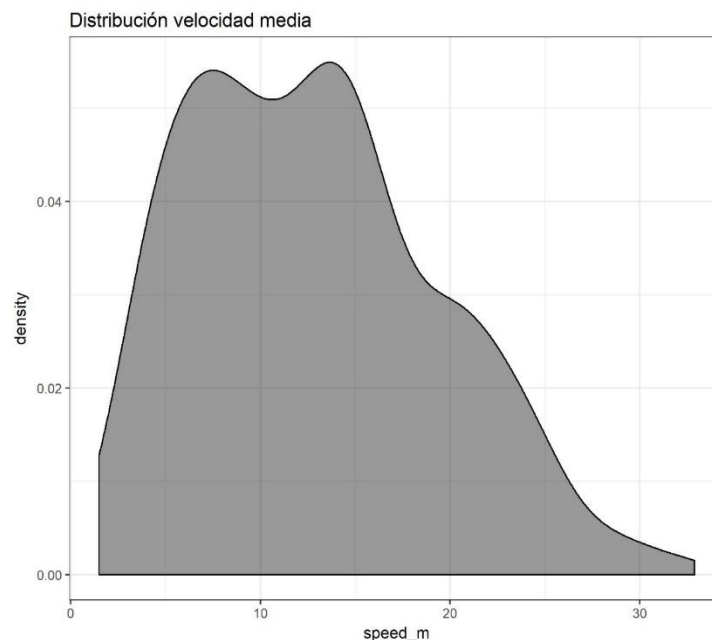


Figura 3.12: Distribución de la velocidad media

La velocidad media de las trayectorias presenta una media de 12.8 km/h y una mediana muy similar de 12.6 km/h. Si nos centramos en analizar todas las velocidades en una misma trayectoria, las gaviotas de Audouin pueden llegar a volar hasta 70 km/h. (En el anexo se adjunta el gráfico de la distribución de velocidad máxima por trayectoria)

3.5.1 Determinación de la estructura de varianzas y covarianzas

Modelo	Efecto aleatorio	Error residual	AIC
1	Homoscedástico	Homoscedástico	2369.9
2	Homoscedástico	Heteroscedástico por tipo de día	2371.9
3	Varianza entre gaviotas diferente según tipo de día	Homoscedástico	2366.5
4	Varianza entre gaviotas diferente según tipo de día	Heteroscedástico por tipo de día	2367.8

Tabla 3.9: Resultados AIC para determinar la estructura de varianzas y covarianzas

```

> anova(m1,m2)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m1     1  3 2369.956 2381.631 -1181.978
m2     2  4 2371.943 2387.510 -1181.972 1 vs 2 0.01341531 0.9078
> anova(m1,m3)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m1     1  3 2369.956 2381.631 -1181.978
m3     2  5 2366.517 2385.975 -1178.258 1 vs 2 7.439705 0.0242
> anova(m3,m4)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m3     1  5 2366.517 2385.975 -1178.258
m4     2  6 2367.836 2391.186 -1177.918 1 vs 2 0.6809911 0.4092

```

Salida R 3.24: Resultados LRT test para determinar la estructura de varianzas y covarianzas

Para la media de velocidades, según el LRT test y el AIC, la **mejor estructura** de varianzas y covarianzas es **una varianza entre gaviotas diferente según el tipo de día y una varianza residual igual**.

3.5.2 Introducción de efectos fijos

Modelo	Efectos fijos	AIC
5	Tipo de día	2353.4
6	Tipo de día y momento del día	2344.3
7	Interacción de los 2 efectos fijos	2343.7

Tabla 3.10: Resultados AIC de los efectos fijos

```
> anova(m5,m6)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m5     1  6 2353.421 2376.771 -1170.711
m6     2  8 2344.324 2375.457 -1164.162 1 vs 2 13.09718  0.0014
> anova(m6,m7)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m6     1  8 2344.324 2375.457 -1164.162
m7     2 10 2343.702 2382.619 -1161.851 1 vs 2  4.62203  0.0992
```

Salida R 3.25: Resultados LRT test de los efectos fijos

Según el LRT test, **el mejor modelo** para explicar la velocidad media es **el que incluye el tipo y momento del día** (aunque el AIC determina que el mejor modelo es con interacción, pero la diferencia de unidades de los AIC es menor que 1 unidad).

El summary del mejor modelo es el siguiente:


```

> summary(m6)
Linear mixed-effects model fit by REML
Data: velocidad_df
      AIC      BIC    logLik
2339.942 2370.987 -1161.971

Random effects:
Formula: ~1 + tipo_dia | id
Structure: General positive-definite, Log-Cholesky parametrization
              StdDev  Corr
(Intercept)   1.141835 (Intr)
tipo_dialaborable 1.694788 0.849
Residual       5.764303

Fixed effects: speed_m ~ tipo_dia + momento
              Value Std.Error DF   t-value p-value
(Intercept)  10.606968 0.8100135 323  13.094804  0.0000
tipo_dialaborable  3.787326 0.8507819 323   4.451583  0.0000
momentotarde    -2.547816 0.7551030 323  -3.374131  0.0008
momentonoche    -0.164700 0.8069285 323  -0.204108  0.8384
Correlation:
              (Intr) tp_dlb mmnttr
tipo_dialaborable -0.634
momentotarde      -0.333 -0.101
momentonoche      -0.330 -0.082  0.438

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3      Max
-2.3987379 -0.7091475 -0.1069500  0.6535316  3.9837406

Number of Observations: 362
Number of Groups: 36

```

Salida R 3.26: Summary del mejor modelo

```

> anova(m6)
              numDF denDF  F-value p-value
(Intercept)      1   323  593.6318 <.0001
tipo_dia          1   323  17.6081  <.0001
momento           2   323   6.6958  0.0014

```

Salida R 3.27: Significación de los efectos fijos

Todos los coeficientes del modelo son estadísticamente significativos. Respecto una trayectoria en día festivo, la velocidad media incrementa en 3.8 en los días laborables (IC 95%: 2.11, 5.46). Esta pendiente es aleatoria con una desviación de 1.64 por el efecto aleatorio de la gaviota.

En la tarde, la velocidad en media disminuye en 2.54 respecto a la mañana. El decremento no es tan brusco en una trayectoria de noche que solo disminuye en -0.17 respecto una trayectoria por la mañana (IC 95%: -1.7, 1.42).

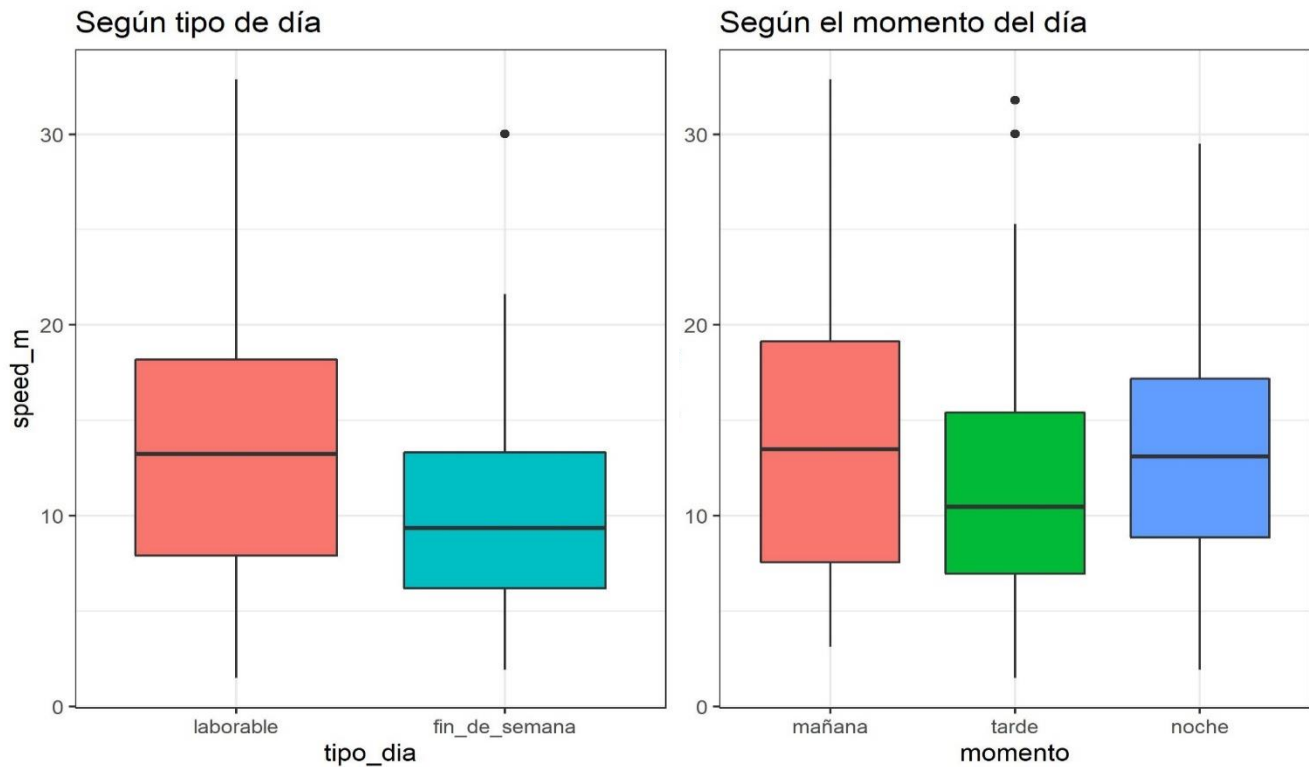


Figura 3.13: Velocidad media según tipo y momento del día

En cuanto a la validación del modelo, se observa que los residuos obtenidos con el mejor modelo se comportan como una distribución normal de media 0. Destaca una sola observación cuyo residuo es 4. Se trata de la trayectoria 5107942V15 que presenta una velocidad media de 30 km/h, muy superior al resto de observaciones.

Se ajusta el mismo modelo sin esta trayectoria: el AIC disminuye en 23 unidades, los coeficientes siguen siendo significativos pero el valor cambia. Se adjunta en el anexo la salida del nuevo modelo.

3.5.3 Repetibilidad de la velocidad media

El ICC de la velocidad media es 0.12 (IC 95%: 0.03, 0.2) lo que indica una repetibilidad baja pero diferente de 0 como indica el test LRT.

```
> anova(m_nomixt, m_mixt)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m_nomixt   1   2 2387.790 2395.573 -1191.895
m_mixt     2   3 2369.956 2381.631 -1181.978 1 vs 2 19.83328 <.0001
```

Salida R 3.28: Resultado test de significación ICC

Si se desglosa por tipo de día, el ICC para los días laborables es de 0.15 (IC 95%: 0.05, 0.26) y el de los días festivos es 0 pero al igual que pasa con las otras repetibilidades, no hay diferencias significativas ($p\text{-valor}=0.23>0.05$).

```
> sum(abs(f.perms) > abs(dif_original))/nperm # two-tailed test  
[1] 0.235
```

Salida R 3.29: Resultado test de permutaciones

3.6 Análisis de las desviaciones de velocidad

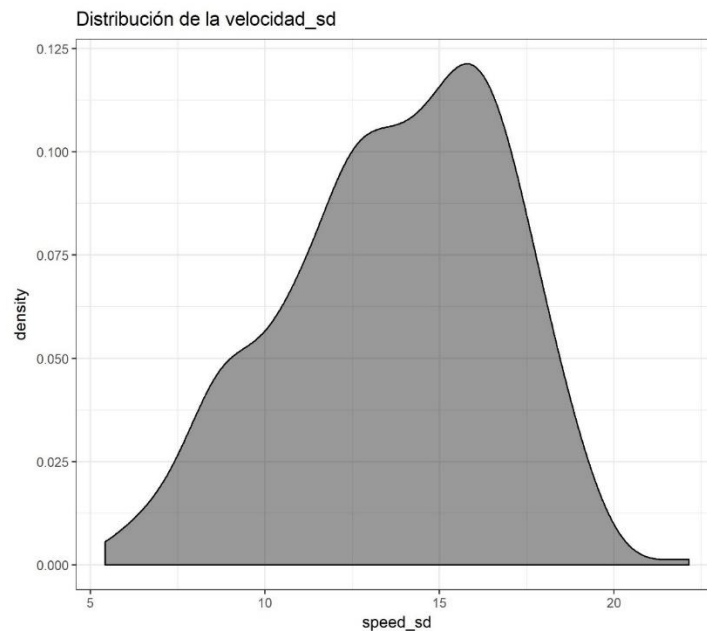


Figura 3.14: Distribución de las desviaciones de velocidad

La desviación media en las trayectorias presenta una media de 13.7 km/h y una mediana de 12.6 km/h. Se trata de una desviación media importante pero no es nada extraño ya que las gaviotas a lo largo del vuelo van parando a comer en los arrozales o en el mar y eso explica la gran variabilidad en sus velocidades.

3.6.1 Determinación de la estructura de varianzas y covarianzas

Modelo	Efecto aleatorio	Error residual	AIC
1	Homoscedástico	Homoscedástico	1812.9
2	Homoscedástico	Heteroscedástico por tipo de día	1814.9
3	Varianza entre gaviotas diferente según tipo de día	Homoscedástico	1816.9
4	Varianza entre gaviotas diferente según tipo de día	Heteroscedástico por tipo de día	1818.9

Tabla 3.11: Resultados AIC para determinar la estructura de varianzas y covarianzas

Ambos criterios concluyen que la **mejor estructura** es la de un **efecto aleatorio homoscedástico y un error residual también homoscedástico**.

```
> anova(m1,m2)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 1812.987 1824.662 -903.4935
m2     2  4 1814.918 1830.485 -903.4590 1 vs 2 0.06893695 0.7929
> anova(m1,m3)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 1812.987 1824.662 -903.4935
m3     2  5 1816.977 1836.435 -903.4883 1 vs 2 0.01044156 0.9948
> anova(m1,m4)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m1     1  3 1812.987 1824.662 -903.4935
m4     2  6 1818.918 1842.268 -903.4590 1 vs 2 0.06894173 0.9953
```

Salida R 3.30: Resultados LRT test para determinar la estructura de varianzas y covarianzas

3.6.2 Introducción de efectos fijos

Modelo	Efectos fijos	AIC
5	Tipo de día	1805.3
6	Tipo de día y momento del día	1787.4
7	Interacción de los 2 efectos fijos	1789.1

Tabla 3.12: Resultados AIC de los efectos fijos

El **mejor modelo** para explicar las desviaciones de velocidad en las trayectorias es aquel que **incluye el tipo de día y el momento del día, sin interacción.**

```
> anova(m5,m6)
Model df      AIC      BIC    logLik  Test L.Ratio p-value
m5     1   4 1805.355 1820.922 -898.6776
m6     2   6 1787.435 1810.785 -887.7176 1 vs 2 21.9201 <.0001
> anova(m6,m7)
Model df      AIC      BIC    logLik  Test L.Ratio p-value
m6     1   6 1787.435 1810.785 -887.7176
m7     2   8 1789.096 1820.229 -886.5479 1 vs 2 2.339453 0.3105
```

Salida R 3.31: Resultados LRT test de los efectos fijos

El summary del modelo es el siguiente:

```

> summary(m6)
Linear mixed-effects model fit by REML
Data: velocidad_df
      AIC      BIC    logLik
1788.785 1812.068 -888.3924

Random effects:
Formula: ~1 | id
(Intercept) Residual
StdDev:      1.570913 2.630687

Fixed effects: speed_sd ~ tipo_dia + momento
              Value Std.Error DF   t-value p-value
(Intercept)  13.147550 0.4540421 323  28.956677 0.0000
tipo_dialaborable  1.288664 0.3682255 323   3.499659 0.0005
momentotarde    -1.379967 0.3506340 323  -3.935633 0.0001
momentonoche      0.267521 0.3746298 323   0.714096 0.4757
Correlation:
              (Intr) tp_dlb mmnttr
tipo_dialaborable -0.625
momentotarde      -0.281 -0.103
momentonoche      -0.277 -0.088  0.444

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3      Max
-2.78952401 -0.58794089  0.03457084  0.68261165  2.27712043

Number of Observations: 362
Number of Groups: 36

```

Salida R 3.32: Summary del mejor modelo

```

> anova(m6)
              numDF denDF   F-value p-value
(Intercept)      1    323  2070.8833 <.0001
tipo_dia          1    323   10.5268  0.0013
momento           2    323   11.5109 <.0001

```

Salida R 3.33: Significación de los efectos fijos

Las desviaciones de velocidad son mayores en días laborables (aumentan en media un 1.28 respecto un día festivo. IC 95%: 0.56, 2.01) La aleatoriedad de la pendiente es en media de 1.57.

En cuanto al momento, es a la tarde donde las desviaciones de velocidad disminuyen respecto a la mañana (coeficiente de -1.37 y el IC 95%: -2, -0.69). En la noche, en cambio hay mayor desviación respecto la mañana, pero el cambio es solo de 0.26.

En cuanto la validación del modelo, los coeficientes son todos significativos, los errores estandarizados son menores que 3 en valor absoluto y de media 0. El Q-Q plot indica normalidad de los errores del modelo.

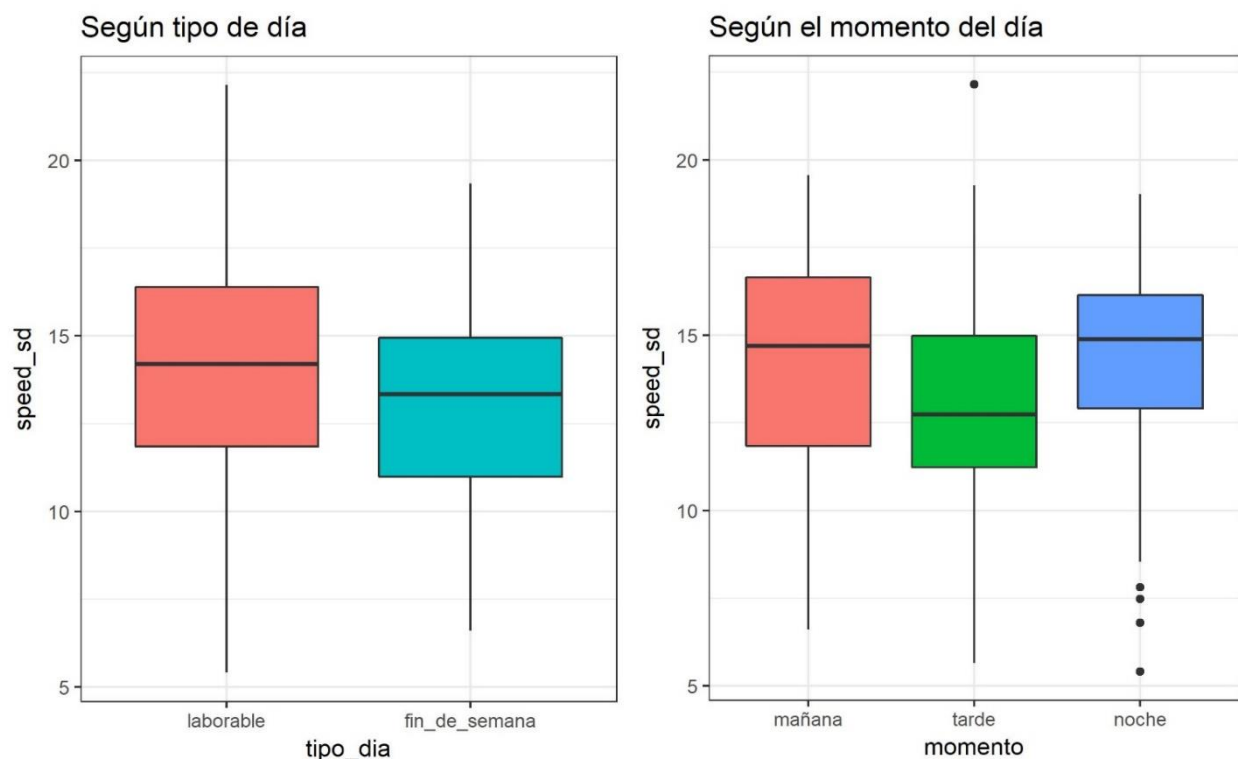


Figura 3.15: Desviaciones de la velocidad según tipo y momento del día

3.6.3 Repetibilidad de las desviaciones de velocidad

El ICC de este índice vale 0.21 (IC 95%: 0.09, 0.32), lo que indica presencia de repetibilidad, pero baja. Esta repetibilidad se mantiene para los días laborables (0.22) pero en los festivos, la repetibilidad baja hasta 0.07 (igual que la repetibilidad de la velocidad media que era nula en los festivos). Sin embargo, no se aprecian diferencias significativas según el tipo de día ($p\text{-valor}=0.34>0.05$).

```
> sum(abs(f.perms) > abs(dif_original))/nperm # two-tailed test
[1] 0.345
```

Salida R 3.34: Resultado test de permutaciones

3.7 Análisis de la aceleración media

A partir de la distribución de la aceleración media, se detecta un valor atípico que se decide eliminar para la modelización (se trata de la trayectoria 5107944V01 con una aceleración media de 79).

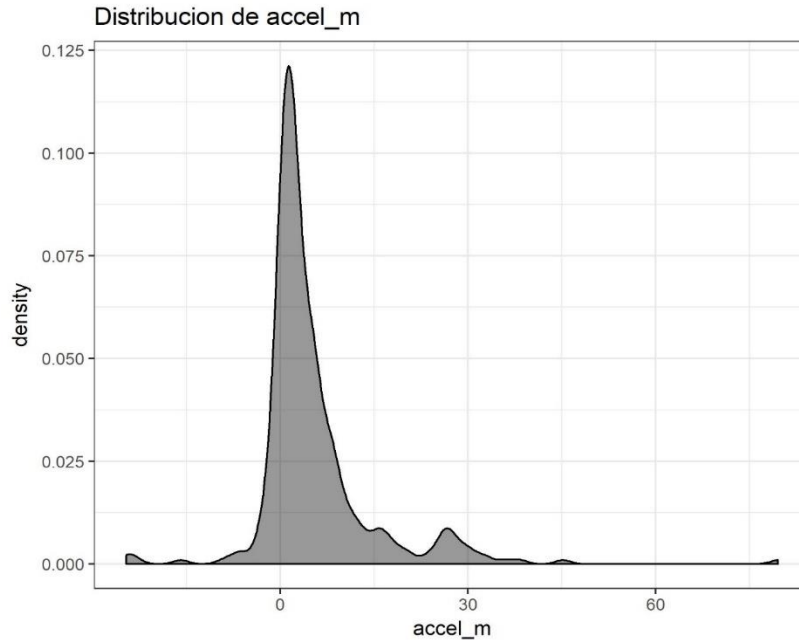


Figura 3.16: Distribución de la aceleración media

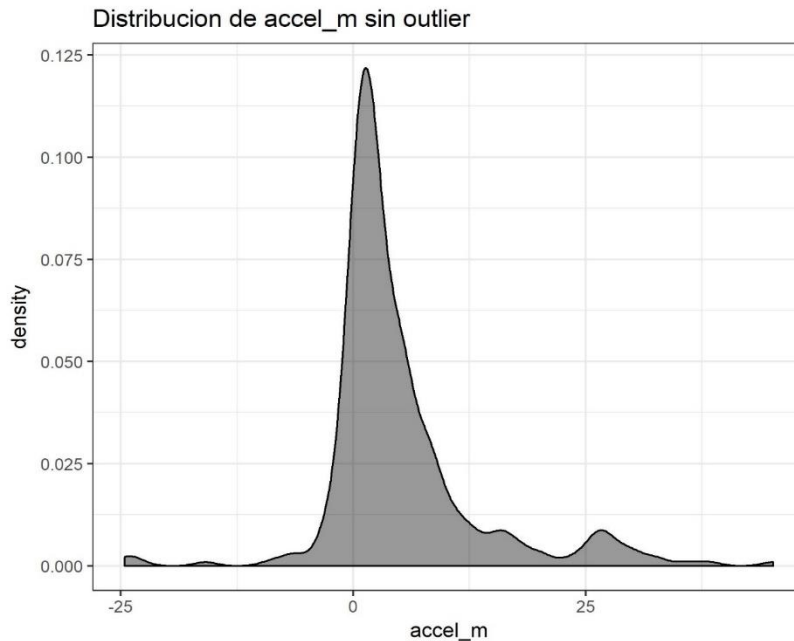


Figura 3.17: Distribución de la aceleración media sin el dato atípico

3.7.1 Determinación de la estructura de varianzas y covarianzas

Modelo	Efecto aleatorio	Error residual	AIC
1	Homoscedástico	Homoscedástico	2547.7
2	Homoscedástico	Heteroscedástico por tipo de día	2549.7
3	Varianza entre gaviotas diferente según tipo de día	Homoscedástico	2551.7
4	Varianza entre gaviotas diferente según tipo de día	Heteroscedástico por tipo de día	2553.6

Tabla 3.13: Resultados AIC para determinar la estructura de varianzas y covarianzas

```

> anova(m1,m2)
  Model df      AIC      BIC    logLik  Test    L.Ratio p-value
m1     1  3 2547.706 2559.373 -1270.853
m2     2  4 2549.698 2565.253 -1270.849 1 vs 2 0.008281963 0.9275
> anova(m1,m3)
  Model df      AIC      BIC    logLik  Test    L.Ratio p-value
m1     1  3 2547.706 2559.373 -1270.853
m3     2  5 2551.703 2571.147 -1270.851 1 vs 2 0.003174323 0.9984
> anova(m1,m4)
  Model df      AIC      BIC    logLik  Test    L.Ratio p-value
m1     1  3 2547.706 2559.373 -1270.853
m4     2  6 2553.696 2577.029 -1270.848 1 vs 2 0.01050339 0.9997

```

Salida R 3.35: Resultados LRT test para determinar la estructura de varianzas y covarianzas

Al igual que con la desviación de la velocidad, la **mejor estructura es un efecto aleatorio y error residual homoscedásticos**. Ambos criterios concluyen lo mismo.

3.7.2 Introducción de efectos fijos

Modelo	Efectos fijos	AIC
5	Tipo de día	2549
6	Tipo de día y momento del día	2551.5
7	Interacción de los 2 efectos fijos	2551.8

Tabla 3.14: Resultados AIC de los efectos fijos

El **mejor modelo** para ambos criterios es **el que incluye solo el tipo de día**. La información sobre el momento del día de la trayectoria no es significativa a la hora de explicar la aceleración media.

```
> anova(m5,m6)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m5     1  4 2633.952 2649.519 -1312.976
m6     2  6 2635.159 2658.508 -1311.579 1 vs 2 2.793934 0.2473
> anova(m5,m7)
  Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m5     1  4 2633.952 2649.519 -1312.976
m7     2  8 2636.947 2668.080 -1310.473 1 vs 2 5.005574 0.2867
```

Salida R 3.36: Resultados LRT test de los efectos fijos

A continuación, el summary del mejor modelo:

```

> summary(m5)
Linear mixed-effects model fit by REML
Data: aceleracion_df2
      AIC      BIC    logLik
2546.254 2561.787 -1269.127

Random effects:
Formula: ~1 | id
      (Intercept) Residual
StdDev:    2.399496  7.94381

Fixed effects: accel_m ~ tipo_dia
              Value Std.Error DF   t-value p-value
(Intercept)  3.928268  1.084772 324  3.621283  0.0003
tipo_dialaborable 0.871646  1.098154 324  0.793738  0.4279
Correlation:
              (Intr)
tipo_dialaborable -0.834

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3      Max
-3.6701854 -0.4528289 -0.2288081  0.1785210  4.9414329

Number of Observations: 361
Number of Groups: 36

```

Salida R 3.37: Summary del mejor modelo

```

> anova(m5)
              numDF denDF  F-value p-value
(Intercept)      1    324 60.35065  <.0001
tipo_dia          1    324  0.63002  0.4279

```

Salida R 3.38: Significación de los efectos fijos

El modelo seleccionado no es válido. El tipo de día no sale como variable significativa para explicar la aceleración media ($p\text{-valor}=0.42>0.05$). Si se observa la distribución según tipo de día y momento del día apenas hay diferencias en la aceleración media.

Los errores no están centrados en 0 y a pesar de eliminar el valor atípico aún hay residuos superiores a 4.

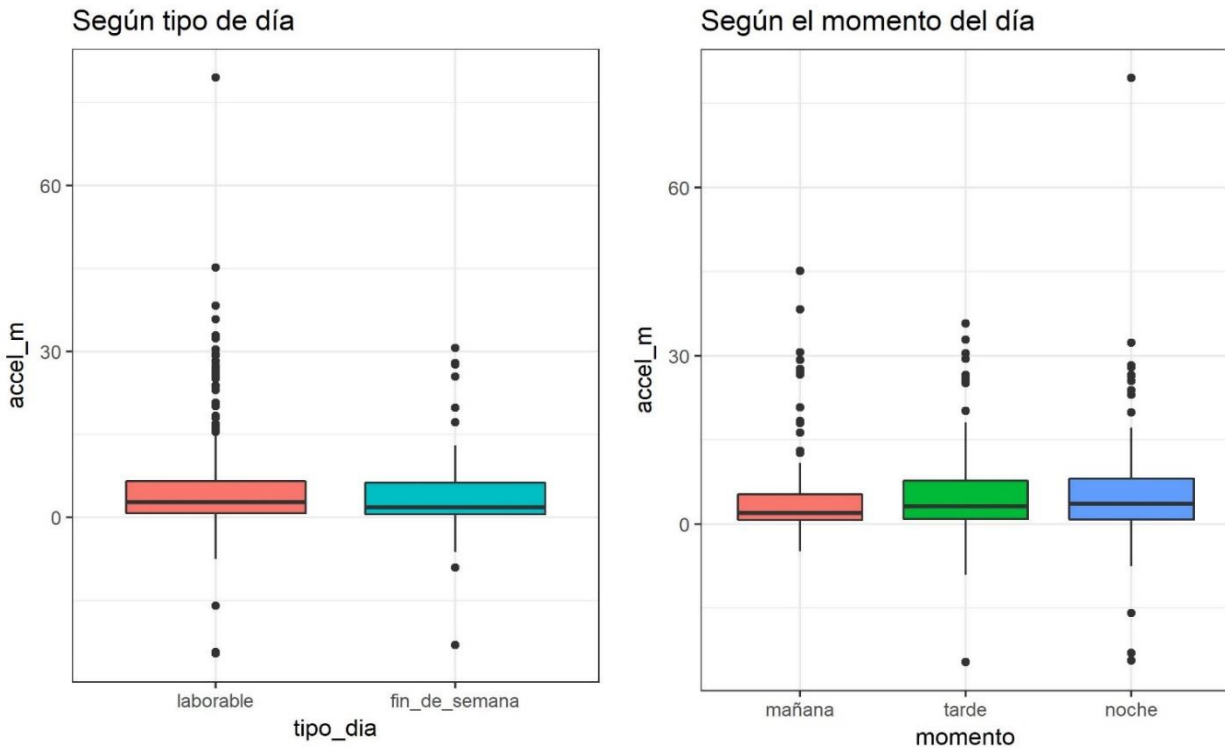


Figura 3.18: Aceleración media según tipo de día y según momento del día

3.7.3 Repetibilidad del índice

El ICC para la aceleración media vale **0.04** (IC 95%: 0, 0.10) pero diferente de 0 (p-valor=0.03<0.05).

```
> anova(m_nomixt,m_mixt)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m_nomixt   1   2 2635.227 2643.010 -1315.613
m_mixt     2   3 2632.802 2644.477 -1313.401 1 vs 2  4.424613  0.0354
```

Salida R 3.39: Resultado test de significación ICC

El resultado no mejora si se separa por tipo de día: ICC para los festivos es de 0 y el ICC para los días laborables;0.04. Los dos ICC no presentan diferencias significativas según el test de permutaciones (p-valor=0.11>0.05).

```
> sum(abs(f.perms) > abs(dif_original))/nperm # two-tailed test
[1] 0.118
```

Salida R 3.40: Resultado test de permutaciones

3.8 Análisis de las desviaciones de aceleración

Se decide aplicar el logaritmo a este índice:

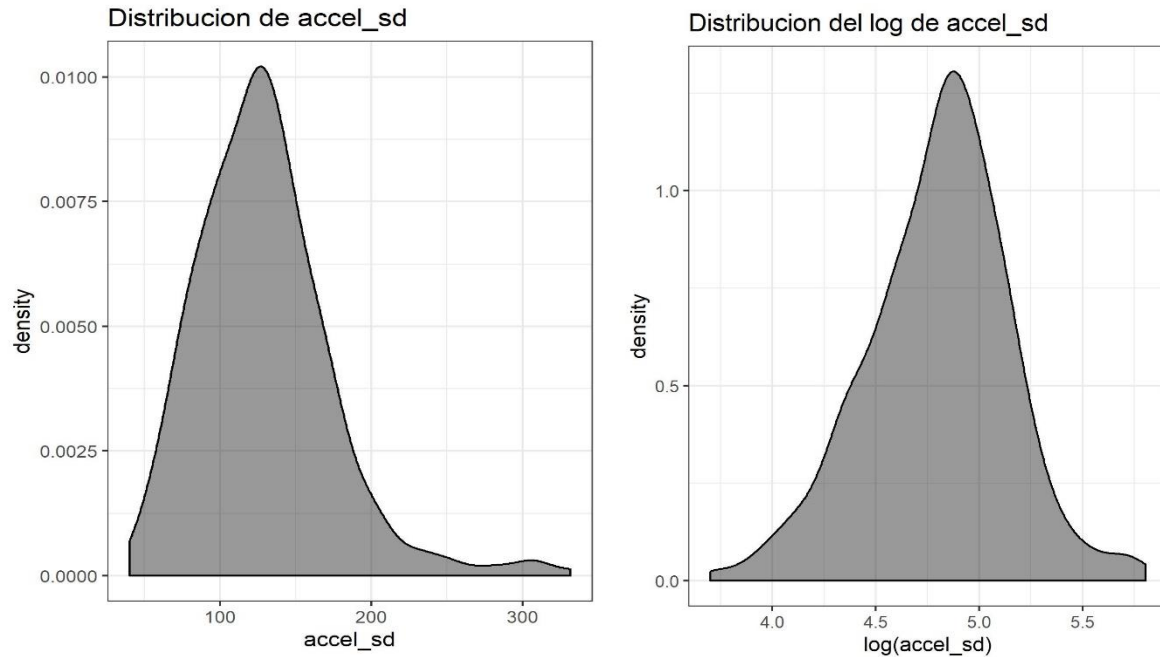


Figura 3.19: Distribución de las desviaciones de aceleración y con logaritmo

3.8.1 Determinación de la estructura de varianzas y covarianzas

Modelo	Efecto aleatorio	Error residual	AIC
1	Homoscedástico	Homoscedástico	236.6
2	Homoscedástico	Heteroscedástico por tipo de día	231
3	Varianza entre gaviotas diferente según tipo de día	Homoscedástico	235.7
4	Varianza entre gaviotas diferente según tipo de día	Heteroscedástico por tipo de día	234.9

Tabla 3.15: Resultados AIC para determinar la estructura de varianzas y covarianzas

Ambos criterios concluyen que **la mejor estructura** de varianza es un **efecto aleatorio homoscedástico y una varianza residual diferente según el tipo de día**.

```
> anova(m1,m2)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m1     1  3 236.6224 248.2973 -115.3112
m2     2  4 231.0503 246.6169 -111.5252 1 vs 2 7.572036 0.0059
> anova(m2,m3)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m2     1  4 231.0503 246.6169 -111.5252
m3     2  5 235.7413 255.1995 -112.8706 1 vs 2 2.690965 0.1009
> anova(m2,m4)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m2     1  4 231.0503 246.6169 -111.5252
m4     2  6 234.8765 258.2264 -111.4383 1 vs 2 0.1738047 0.9168
```

Salida R 3.41: Resultados LRT test para determinar la estructura de varianzas y covarianzas

3.8.2 Introducción de efectos fijos

Modelo	Efectos fijos	AIC
5	Tipo de día	225.1
6	Tipo de día y momento del día	226.9
7	Interacción de los 2 efectos fijos	230.6

Tabla 3.16: Resultados AIC de los efectos fijos

```
> anova(m5,m6)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m5     1  5 225.1057 244.5639 -107.5528
m6     2  7 226.9439 254.1854 -106.4720 1 vs 2 2.161742 0.3393
> anova(m5,m7)
  Model df      AIC      BIC    logLik   Test  L.Ratio p-value
m5     1  5 225.1057 244.5639 -107.5528
m7     2  9 230.5951 265.6199 -106.2976 1 vs 2 2.510569 0.6427
```

Salida R 3.42: Resultados LRT test de los efectos fijos

El **mejor modelo**, tanto para el AIC como el LRT test es el que incluye **solo el tipo de día como variable explicativa**.

El summary del mejor modelo:

```
> summary(m5)
Linear mixed-effects model fit by REML
Data: aceleracion_df
      AIC      BIC    logLik
234.5738 254.0043 -112.2869

Random effects:
Formula: ~1 | id
      (Intercept)  Residual
StdDev:   0.1340227 0.3706365

Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | tipo_dia
Parameter estimates:
fin_de_semana      laborable
      1.0000000      0.8062804
Fixed effects: log(accel_sd) ~ tipo_dia
              Value Std.Error DF t-value p-value
(Intercept)   4.686102 0.05205673 325 90.01914  0.0000
tipo_dialaborable 0.142942 0.04959411 325  2.88223  0.0042
Correlation:
              (Intr)
tipo_dialaborable -0.837

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3      Max
-2.8782396 -0.6093426  0.0464692  0.5611581  3.2362749

Number of Observations: 362
Number of Groups: 36
```

Salida R 3.43: Summary del mejor modelo

```
> anova(m5)
              numDF denDF  F-value p-value
(Intercept)      1    325 28499.734  <.0001
tipo_dia          1    325   8.307  0.0042
```

Salida R 3.44: Significación de los efectos fijos

La variable tipo de día es significativa para explicar las desviaciones de la aceleración. Respecto a los días festivos, la varianza residual en los días laborables es 20 puntos porcentuales menos (en escala proporcional: un 25% inferior).

Es en los días laborables en que hay más desviación en las aceleraciones de las trayectorias (un aumento de 0.15 en el log de la desviación respecto a días festivos. IC 95%: 0.04, 0.24). En escala proporcional, las desviaciones del cambio direccional en las trayectorias de días laborable son un 16% superior a la de los días festivos.

En cuanto a la validación, el modelo presenta unos residuos normales (el qq-plot de los residuos así lo demuestra) y centrados en 0 y menores de 3 en valor absoluto, a excepción de un solo valor.

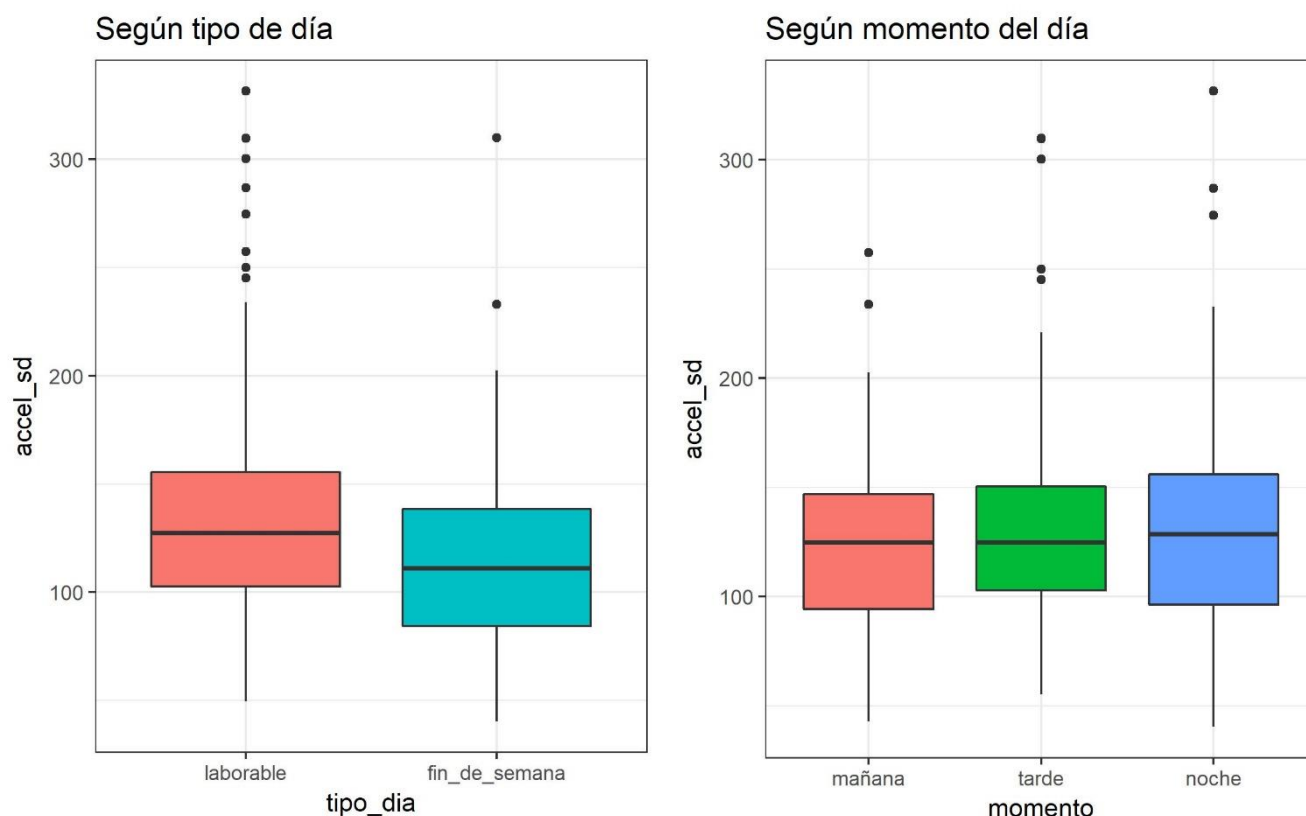


Figura 3.20: Desviaciones de la aceleración según tipo de día y según momento del día

3.8.3 Repetibilidad de las desviaciones de aceleración

El ICC de este índice vale **0.11** (IC 95%: 0.02, 0.2) lo que indica una **repetibilidad bastante baja pero diferente de 0** como concluye el test LRT (pvalor<0.05).

```
> anova(m_nomixt, m_mixt)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
m_nomixt   1  2 3783.882 3791.665 -1889.941
m_mixt     2  3 3770.444 3782.119 -1882.222 1 vs 2 15.43836 1e-04
```

Salida R 3.45: Resultado test de significación ICC

Si se desglosa por tipo de día, el ICC para los días laborables es de 0.14 mientras que para los días festivos es menor, con valor de 0.04 (se podría decir nula repetibilidad) aunque en realidad no presentan diferencias estadísticamente significativas como concluye el test de permutaciones (p-valor=0.51>0.05).

```
> sum(abs(f.perms) > abs(dif_original))/nperm # two-tailed test
[1] 0.518
```

Salida R 3.46: Resultado test de permutaciones

CAPÍTULO 4: Análisis de Componentes Principales

Con la información de los 8 índices por cada trayectoria se realiza un análisis de componentes principales para ver la relación de los índices de manera conjunta.

Se han seleccionado como variables suplementarias las variables día de salida y momento del día.

La representación de las trayectorias en el plano factorial de las dos primeras componentes principales es la siguiente:

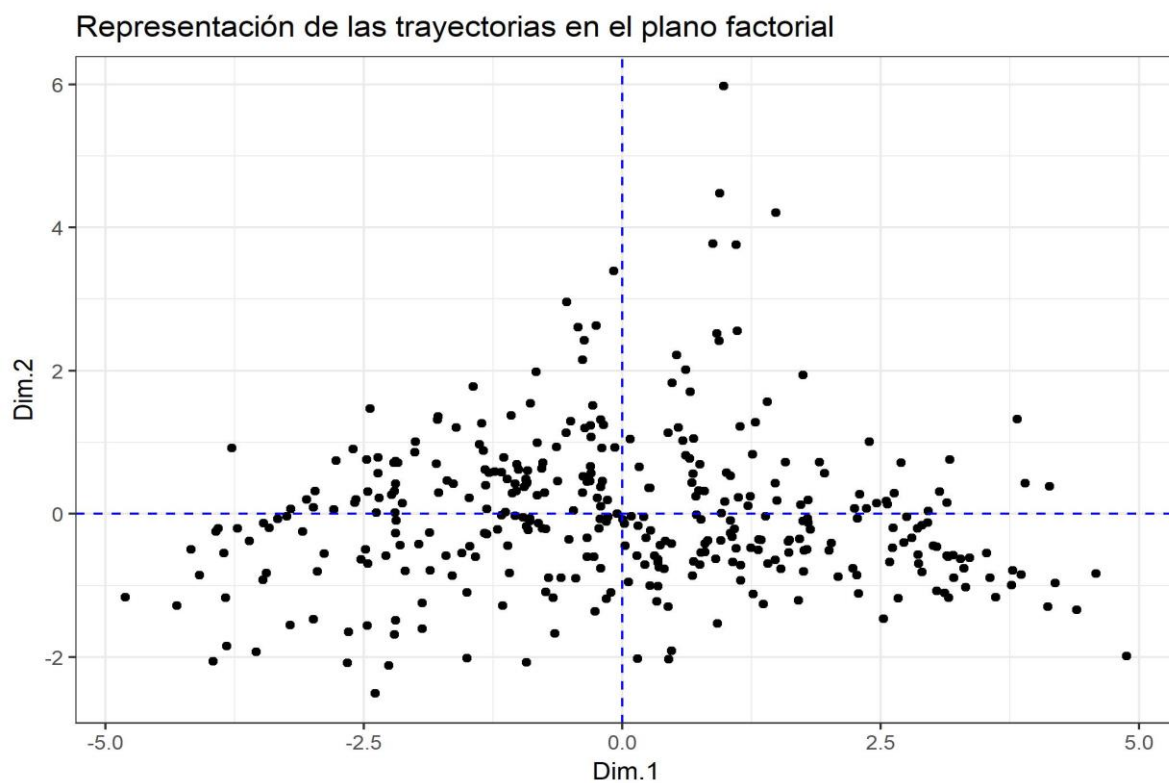


Figura 4.1: Plano factorial de las dos primeras componentes

4.1 Representación de las variables numéricas

Las dos primeras componentes explican más de un 60% de la variabilidad total de los datos siendo la primera componente la que explica casi un 50% de la variabilidad.

A continuación, se adjunta el gráfico con la representación de los índices y su interpretación. Los valores de correlación que se indica en la interpretación son los de la correlación de Pearson.

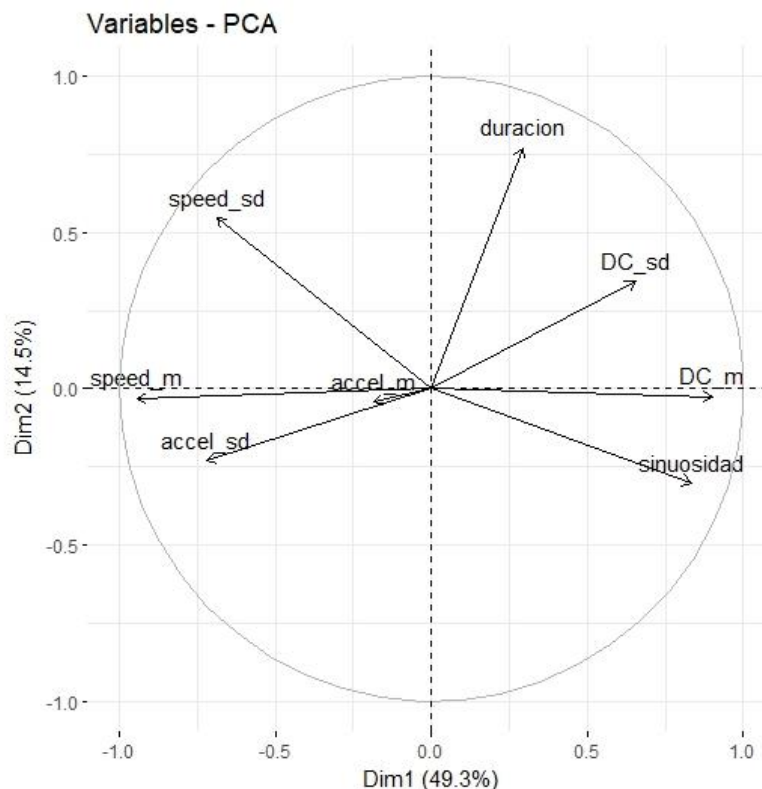


Figura 4.2: Representación índices

El índice de sinuosidad se encuentra correlacionado positivamente con el índice de no linealidad (correlación de 0.7). Estos dos índices, se encuentran correlacionados muy negativamente con la velocidad media (ambas correlaciones son menor de -0.7).

Además, en el caso de la sinuosidad, también se encuentra correlacionada negativamente con la variabilidad en la velocidad de la trayectoria. (correlación negativa de 0.71)

La interpretación es que trayectorias con mucha tortuosidad y no lineales presentan en media una velocidad menor.

Las trayectorias con mayor velocidad se asocian a una mayor variabilidad en las aceleraciones de las trayectorias (correlación de **speed_m** y **aceleracion_sd** es positiva con valor 0.6)

Además, destaca el hecho de que la duración de la trayectoria es la única variable que no está correlacionada con ningún índice (ninguna correlación con los otros índices es mayor que 0.3 en valor absoluto).

Por último, el índice de aceleración media es el único que no está bien representado en el plano factorial.

4.2 Representación conjunta de variables en el plano factorial

Las variables suplementarias también se pueden representar en el plano factorial. A continuación, se adjunta la representación conjunta de los índices con el centro de gravedad de las categorías de día de salida en la trayectoria y momento del día.

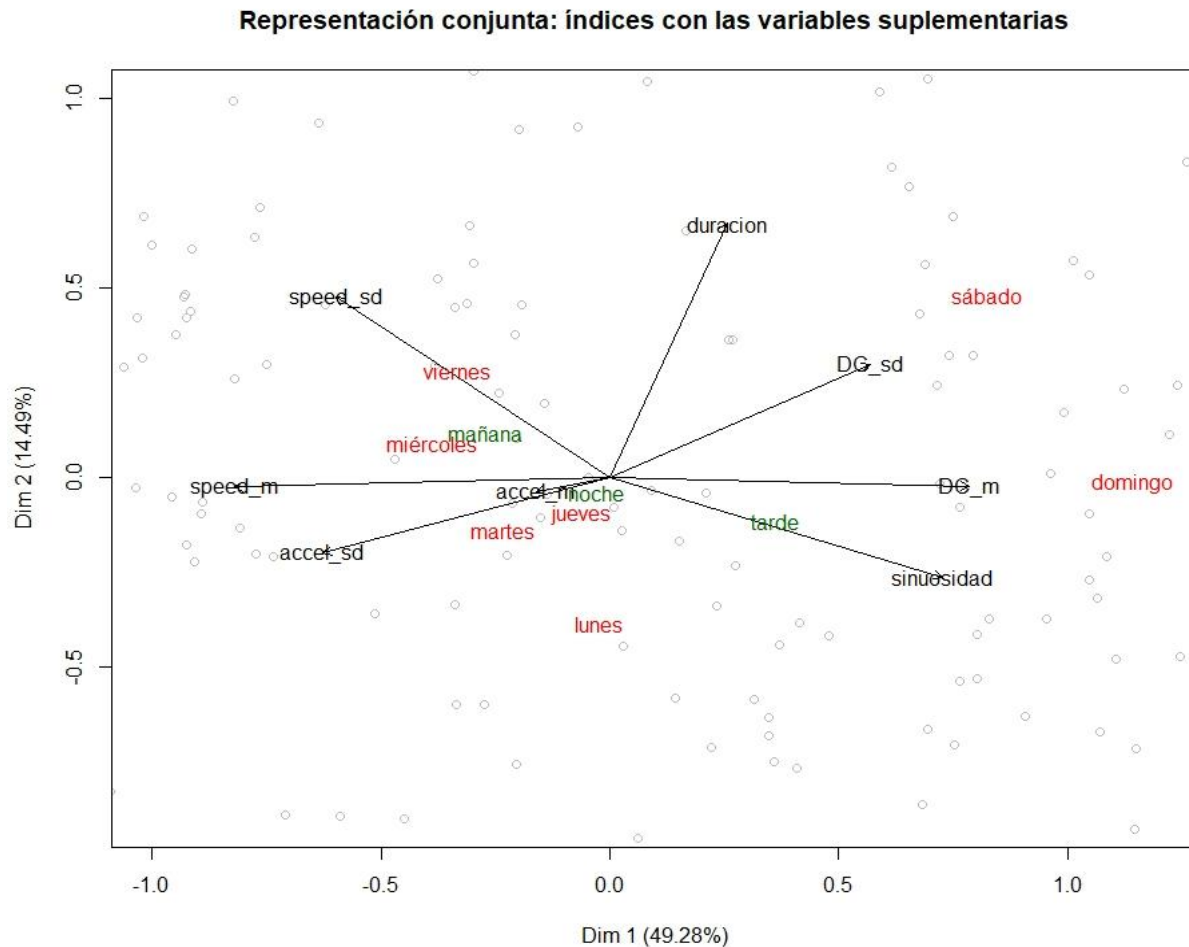


Figura 4.3: Representación conjunta

Las trayectorias que se inician el fin de semana se asocian a mayor duración, sinuosidad, irregularidad y no linealidad (como se pudo comprobar también con los modelos mixtos). Claramente vemos que en el fin de semana las trayectorias son completamente distintas ya que los centros de gravedad de “sábado” y “domingo” están bastante alejados de las demás categorías.

En cambio, los días laborables, las trayectorias tienen mayor media y variabilidad en las velocidades y aceleraciones.

4.3 Interpretación de las dos primeras componentes

Por último, para poder interpretar las dos primeras componentes resultantes del PCA se calcula la contribución de cada índice en cada componente:

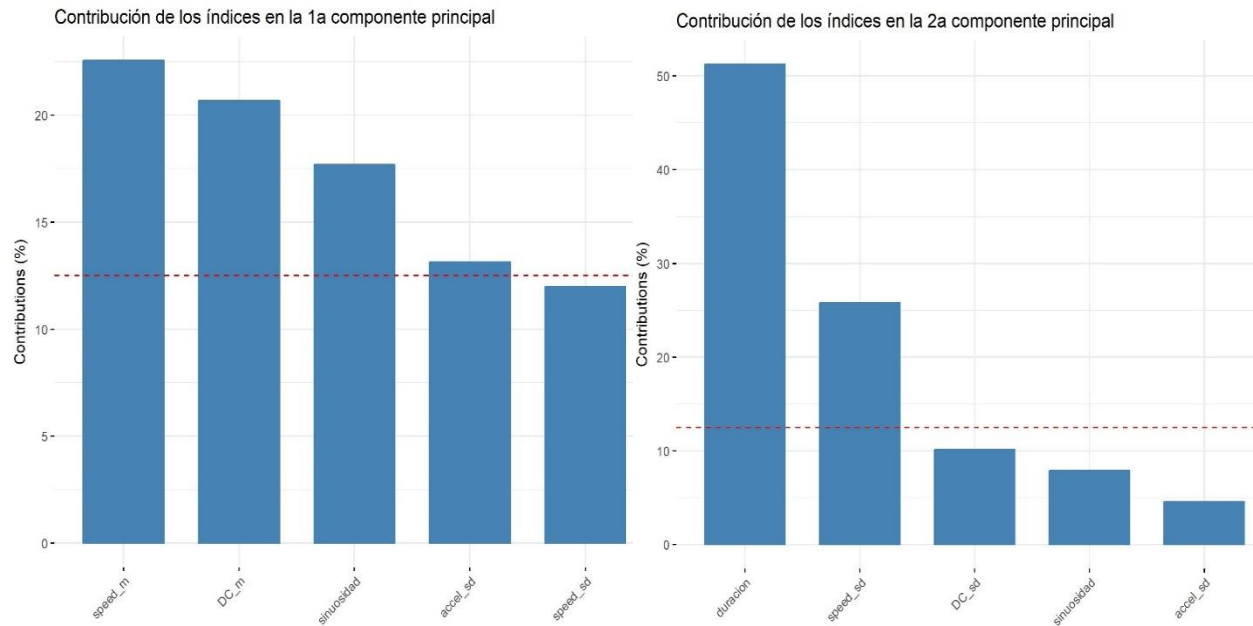


Figura 4.4: Contribución de los índices

La primera variable latente está fuertemente relacionada con la velocidad media, la no linealidad (media del cambio direccional) y la sinusoidalidad. En cambio, los índices de duración, variabilidad de la velocidad contribuyen fuertemente en la creación de la segunda variable latente.

Así pues, la primera componente está más relacionada con la forma que toma la trayectoria (por la media del cambio direccional y la sinusoidalidad) a diferencia de la segunda variable latente que no toma en consideración la forma de la trayectoria sino más bien el tiempo de trayectoria (por la duración que también está ligada con las desviaciones de velocidad)

CAPÍTULO 5: Aplicativo Shiny

En este breve capítulo se explica cómo se ha implementado una aplicación interactiva para visualizar la información que proporcionan las trayectorias de las gaviotas

Justificación de implementar una aplicación

En el ámbito del data science, es importante saber comunicar los resultados de un análisis a un público amplio y es por ello que Shiny es una gran opción ya que permite presentar los resultados de una forma interactiva mediante tablas, gráficos y otras opciones de visualización.

El dashboard final consta de 2 pestañas.

A continuación, se explica detalladamente la información que ofrece cada pestaña.

Pestaña1: Visualización de las trayectorias

El primer apartado permite al usuario de la aplicación elegir una determinada trayectoria mediante dos desplegables: selección de la anilla (identificador gaviota) y el identificador del viaje. Este último depende de la selección de la gaviota ya que no todas hacen el mismo número de viajes por lo que la creación de este desplegable se realiza en el server.

Una vez seleccionada la trayectoria se puede visualizar en el mapa leaflet, así como unas cajas de resultados (valueBox) que notifican el % de duración que vuela por la mañana, tarde, noche o madrugada. Todas las trayectorias de una gaviota comparten el mismo mapa para poder visualizarlas mejor.

Además, la trayectoria está pintada según cada momento del día para que la información sea visualmente atractiva y sepamos en donde estaba en cada uno de los momentos.

Por último, si se clica en el inicio de la trayectoria, aparece información adicional sobre ésta: día de salida, sinuosidad, duración y fechas de inicio y final

El aspecto de esta pestaña se muestra a continuación:

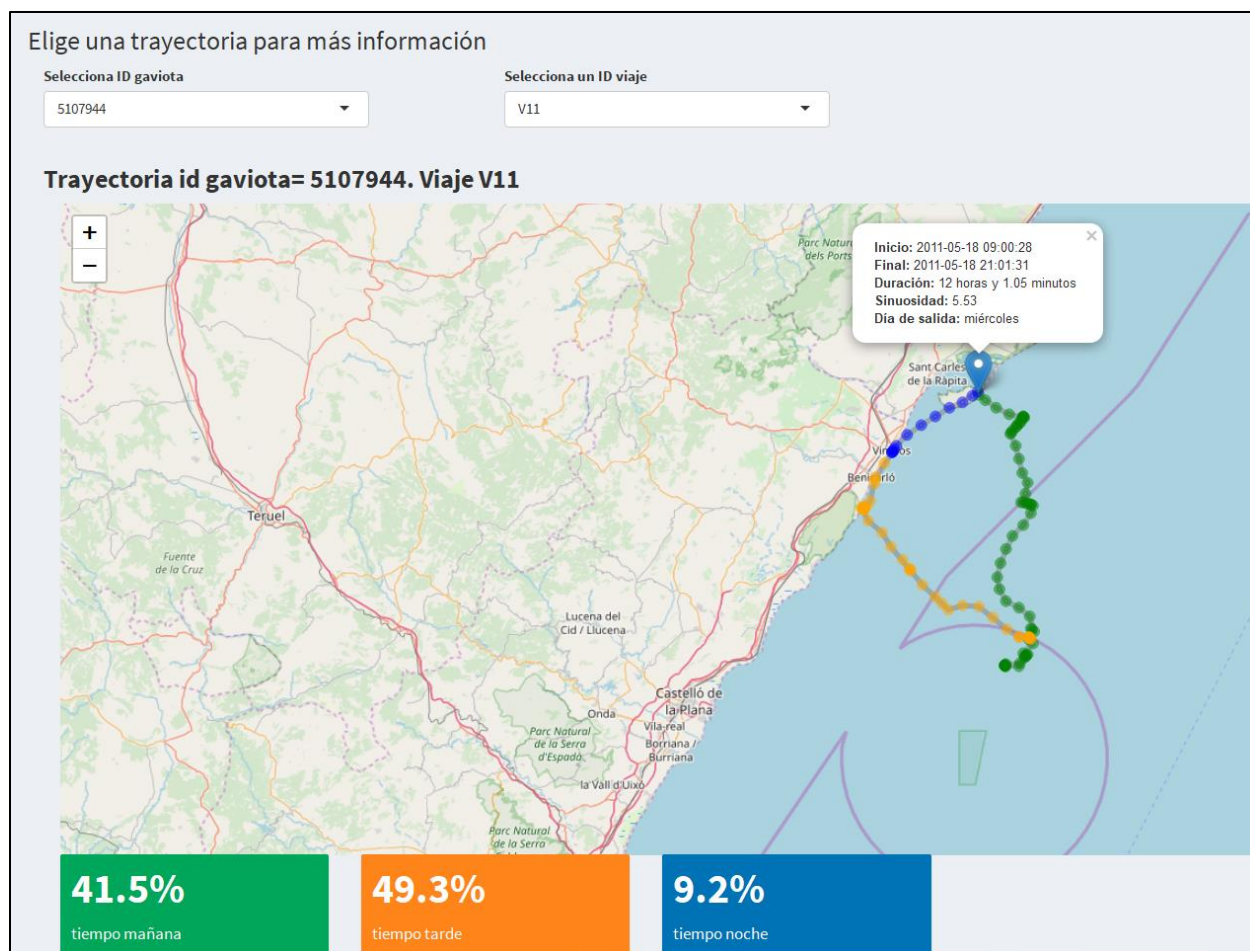


Figura 5.1: Primera pestaña del dashboard

En este ejemplo se muestra la trayectoria 5107944V11 que empieza en la mañana y vuelve al nido por la noche: la trayectoria sucede un 41.5% del tiempo en la mañana, un 49.3% por la tarde y solo un 9.2% por la noche. Además en el popup nos informa que la trayectoria se inició el miércoles y duró 12 horas y 1 minuto.

Pestaña 2: Ranking de trayectorias

En la segunda pestaña, el usuario puede elegir ver hasta un máximo de 10 trayectorias ordenadas de forma ascendente o descendente según su sinuosidad o duración.

De esta manera el usuario se puede hacer una idea de lo diferente que son las trayectorias en cuanto a tortuosidad y duración.

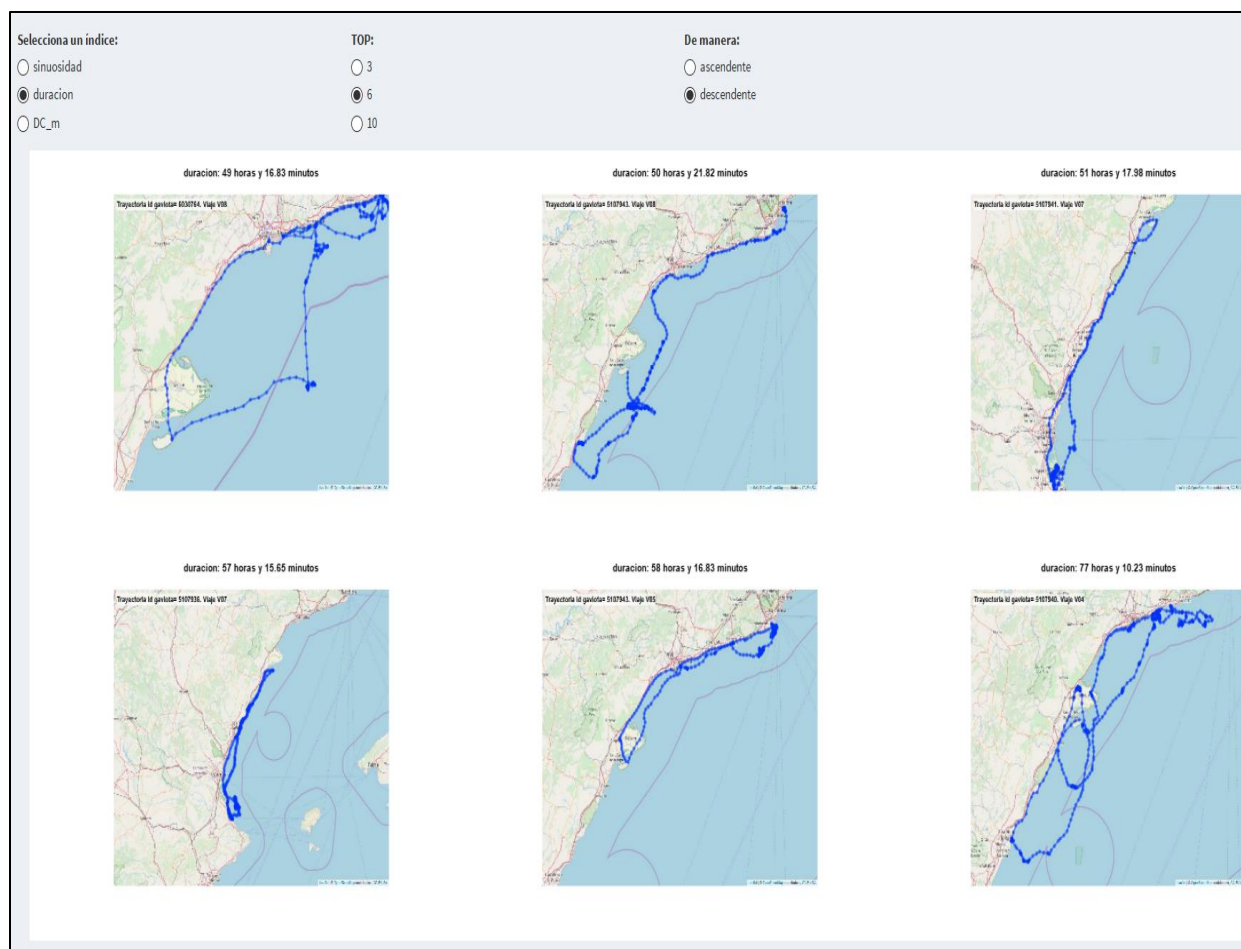


Figura 5.2: Segunda pestaña del dashboard. Visualización de las 6 trayectorias con mayor duración

Conclusiones

Resultados

Claramente hay diferencias significativas en las trayectorias según el tipo de día y momento del día tal y como se observó en los modelos mixtos.

A modo de tabla resumen estas fueron las conclusiones para cada índice:

Índice	Varianza entre	Varianza residual	Factores significativos
Sinuosidad	Diferente	Diferente	Tipo y momento del día
Duración	Igual	Diferente	Tipo y momento del día
Media DC	Diferente	Diferente	Interacción tipo y momento del día
Desviación DC	Diferente	Diferente	Interacción tipo y momento del día
Velocidad media	Diferente	Igual	Tipo y momento del día
Desviación velocidad	Igual	Igual	Tipo y momento del día
Aceleración media	Igual	Igual	Ninguno
Desviación acelerac.	Igual	Diferente	Tipo de día

En todos los modelos, a excepción de los dos de aceleración, ambos factores (tipo y momento del día) son estadísticamente significativos.

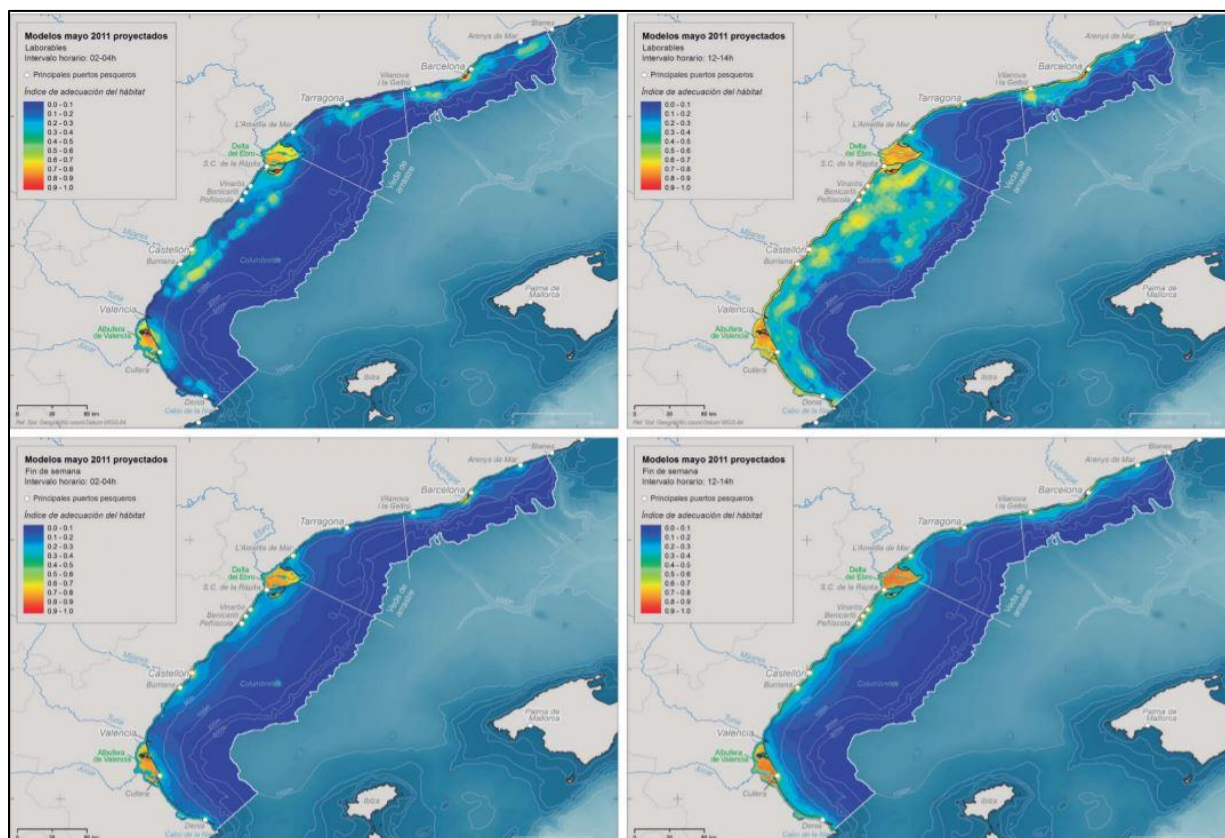
Además, en los índices de media y desviación del cambio direccional de las trayectorias tanto el tipo de día como el momento cuando vuelan y su interacción son variables significativas a la hora de explicar la variabilidad de estos índices.

En cambio, la aceleración media de las trayectorias es el único índice en donde ni el tipo ni momento del día logran ser estadísticamente significativos. También es el único índice mal representado en el análisis de componentes principales.

Centrándose en la interpretación y diferencias según tipo y momento del día, el estudio del SEO Bird Life (*Migración y ecología espacial de la gaviota de Audouin en el Mediterráneo occidental y noroeste africano*) fue de gran ayuda para entender el porqué de esto. El estudio determina que hay diferencias en la ocupación del hábitat según el tipo de día debido a la influencia de la flota pesquera en los días laborables.

Esto puede dar una explicación al hecho de que en los días laborables las trayectorias tengan una media menor en los cambios direccionales, así como en la sinuosidad de las trayectorias ya que las gaviotas van directas a donde tienen que ir a comer. En cambio, el fin de semana realizan un vuelo menos eficiente, en el sentido de más tortuoso y de más duración y con más variabilidad en los cambios de dirección, ya que van buscando alimento.

En este gráfico del estudio mencionado se ve claramente que el fin de semana (segunda fila) las gaviotas pasan más tiempo cerca de la colonia (en la descriptiva de este trabajo se vio que los domingos la distancia máxima al nido era mucho menor que los otros días), en cambio en los días laborables vuelan cerca de los puertos pesqueros.



Respecto a los tres momentos del día, destaca la tarde que es cuando hay más diferencias con la mañana y la noche: la sinuosidad es mayor por las tardes y muy similar entre mañana y noche. En el cambio direccional es a la tarde donde menos diferencias hay según día laborable o no (mientras que en la mañana y en la noche, el cambio direccional en el fin de semana es mayor al de los días laborables). En las velocidades, tanto en media como en variabilidad, la tarde es donde menor media y menos variabilidad hay, mientras que por la mañana y por la noche los valores son mayores que a la tarde, pero parecidos entre sí.

En cuanto a los dos métodos de comparación de modelos, se concluye que en casi todas las comparaciones tanto el AIC como el LRT test concluían lo mismo. Solamente hubo divergencia

en 2 comparaciones: en ver la significación estadística de la interacción entre tipo y momento del día en los índices de sinuosidad y velocidad media.

La repetibilidad de todos los índices es diferente de 0, pero bastante baja, siendo la sinuosidad el índice con mayor repetibilidad con un ICC de 0.3 y la aceleración media es el que menor ICC, con valor de 0.04. Además, el test de permutaciones concluye en todos los índices que no hay diferencias significativas en la repetibilidad según el tipo de día.

El PCA sirvió para corroborar las 2 variables latentes que describen la mecánica del vuelo: la forma de la trayectoria (por la importancia de la sinuosidad y el cambio direccional en la primera componente) y la magnitud del vuelo (por la importancia en la segunda componente de la duración y desviaciones de la velocidad).

Posibles extensiones

Se puede utilizar otros índices para caracterizar la mecánica del vuelo como por ejemplo la autocorrelación direccional que detecta periodicidades, pero es computacionalmente muy intensiva de calcular por lo que se descartó del análisis.

Como se explicó en el capítulo 2, en este trabajo se utiliza las posiciones de GPS que se recogen aproximadamente cada 5 minutos. Cabe aclarar que entre punto y punto la trayectoria del punto no tiene por qué ser recta (es un supuesto de este análisis). Esto es así cuando las gaviotas se alimentan, pero durante los viajes entre colonia y las zonas de alimentación los movimientos son más lineales por lo que el sesgo es prácticamente nulo (Bécares, J.; Arcos, J. M. y Oro, D. 2016)

Dicho esto, se podría mejorar el presente trabajo si se dispusiera de la trayectoria real de cada gaviota.

Bibliografía

Artículos

Bécares, J.; Arcos, J. M. y Oro, D. (2016). *Migración y ecología espacial de la gaviota de Audouin en el Mediterráneo occidental y noroeste africano*. Monografía n.º 1 del programa Migra. SEO/BirdLife. Madrid.

Benhamou, S. (2004). *How to reliably estimate the tortuosity of an animal's path*. Journal of Theoretical Biology, 229(2), 209–220. <https://doi.org/10.1016/j.jtbi.2004.03.016>

Calenge, C.; Dray, S.; & Royer-Carenzi, M. (2009). *The concept of animals' trajectories from a data analysis perspective*. Ecological Informatics, 4(1), 34–41. <https://doi.org/10.1016/j.ecoinf.2008.10.002>

Carrasco, JL, S King, T.; Chinchili, V.; (2009). *The concordance Correlation Coefficient for Repeated Measures Estimated by Variance Components*

McLean, DJ.; Skowron Volponi, MA. (2018). *Trajr: An R package for characterization of animal trajectories*. Ethological Methods; 00:1–9. <https://doi.org/10.1111/eth.12739>.

Ouled-Cheikh, Jazel.; Sanpera, C.; Bécarez, J.; Arcos, JM.; Carrasco, JL., Ramirez, F. (2019). *Spatiotemporal analyses on tracking data reveal fine-scale, daily cycles in seabird fisheries interactions*.

TY - JOUR AU - Kitamura, Tasuku AU - Imafuku, Michio PY - 2015/06/22 SP - T1 - Behavioural mimicry in flight path of Batesian intraspecific polymorphic butterfly *Papilio polytes* VL - 282 DO - 10.1098/rspb.2015.0483

Páginas web (última consulta: 27/12/2019)

<https://shiny.rstudio.com/tutorial/>

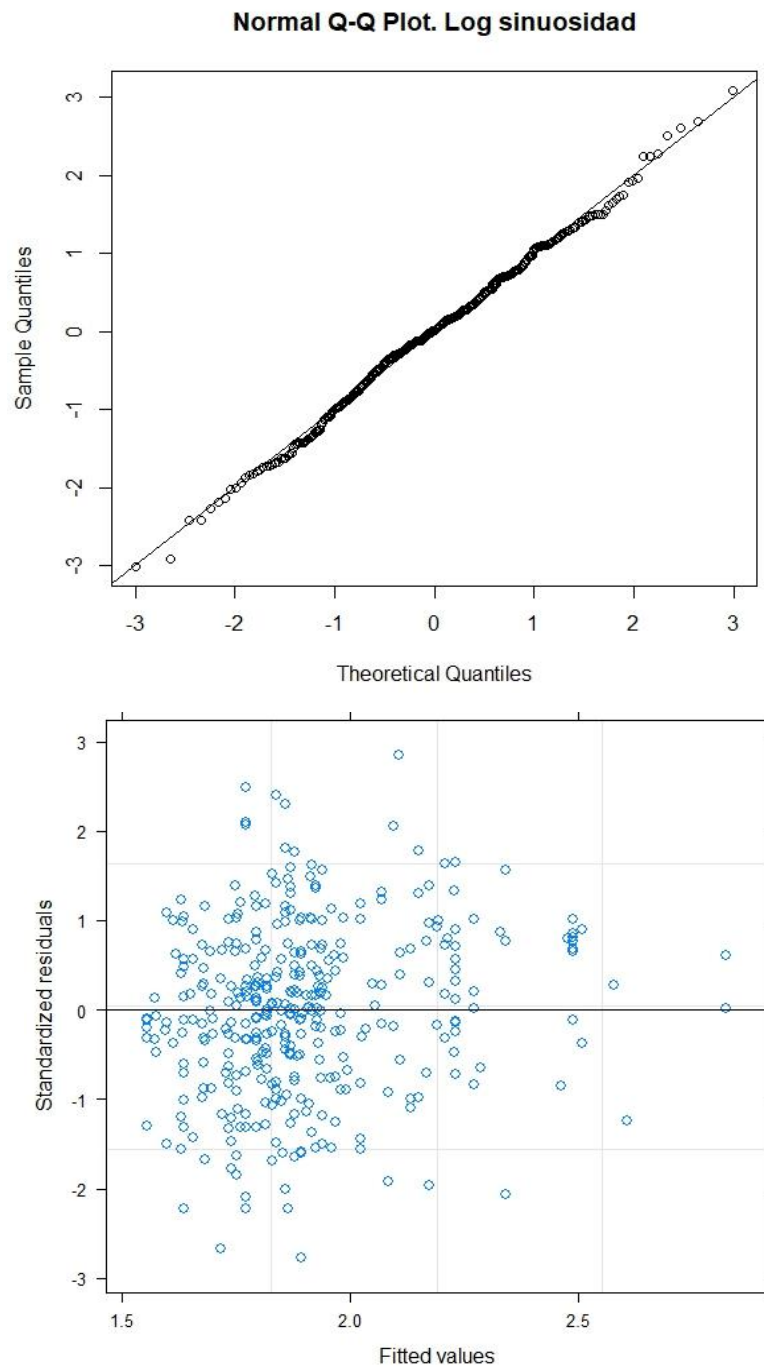
<https://rstudio.github.io/leaflet/>

<https://www.seo.org/ave/gaviota-de-audouin/>

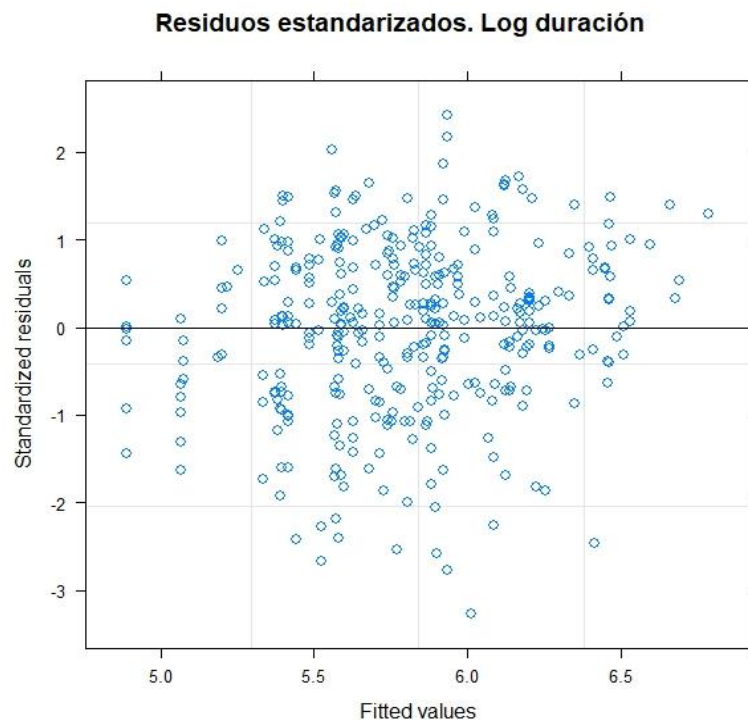
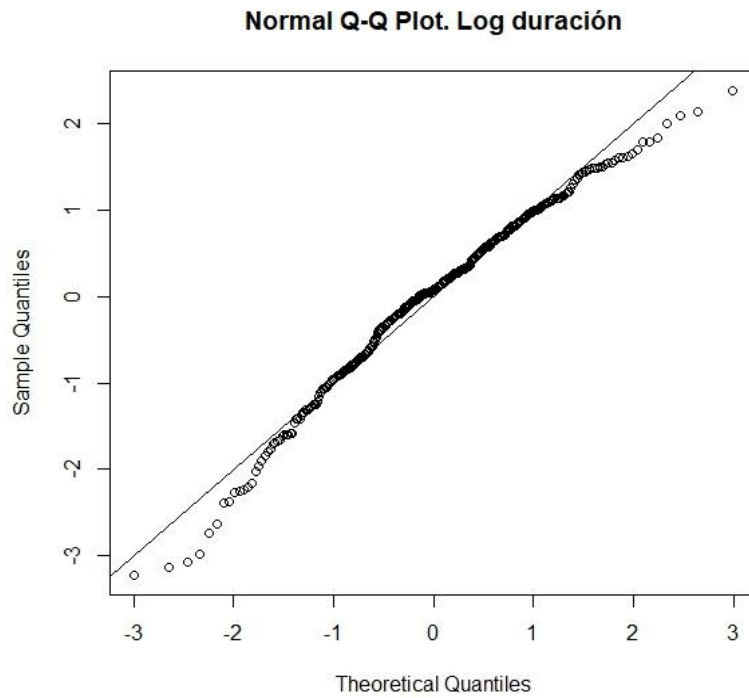
<https://cran.rstudio.com/web/packages/trajr/vignettes/trajr-vignette.html>

Anexo

Validación sinuosidad



Validación duración



```

> summary(m6sin_outlier)
Linear mixed-effects model fit by maximum likelihood
Data: idxs[-115, ]
      AIC      BIC    logLik
978.5744 1005.797 -482.2872

Random effects:
Formula: ~1 | id
      (Intercept) Residual
StdDev:   0.3429163 1.045934

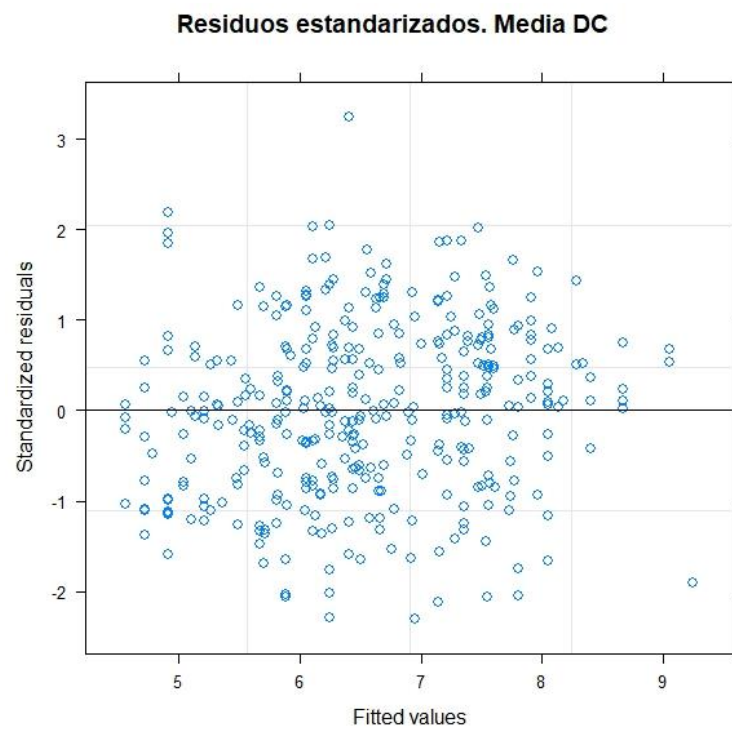
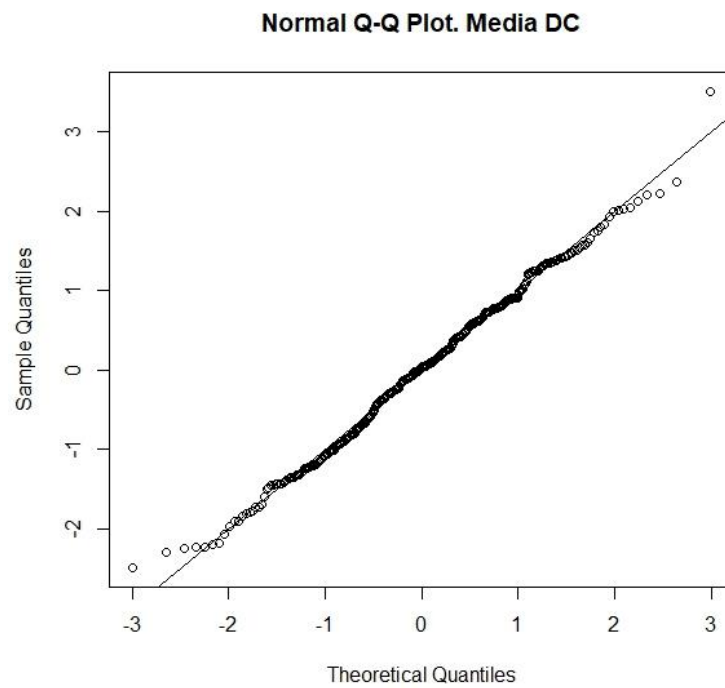
Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | tipo_dia
Parameter estimates:
fin_de_semana    laborable
      1.0000000    0.8107502
Fixed effects: log(duration) ~ tipo_dia + momento
              Value Std.Error DF  t-value p-value
(Intercept)   6.403374 0.1531034 322 41.82385  0.0000
tipo_dialaborable -0.312541 0.1414908 322 -2.20891  0.0279
momentotarde    -0.538973 0.1156202 322 -4.66158  0.0000
momentonoche    -0.365163 0.1233785 322 -2.95970  0.0033
Correlation:
              (Intr) tp_dlb mmnttr
tipo_dialaborable -0.770
momentotarde      -0.274 -0.093
momentonoche      -0.269 -0.077  0.448

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-2.85471958 -0.66809774  0.05918051  0.69793084  2.68906527

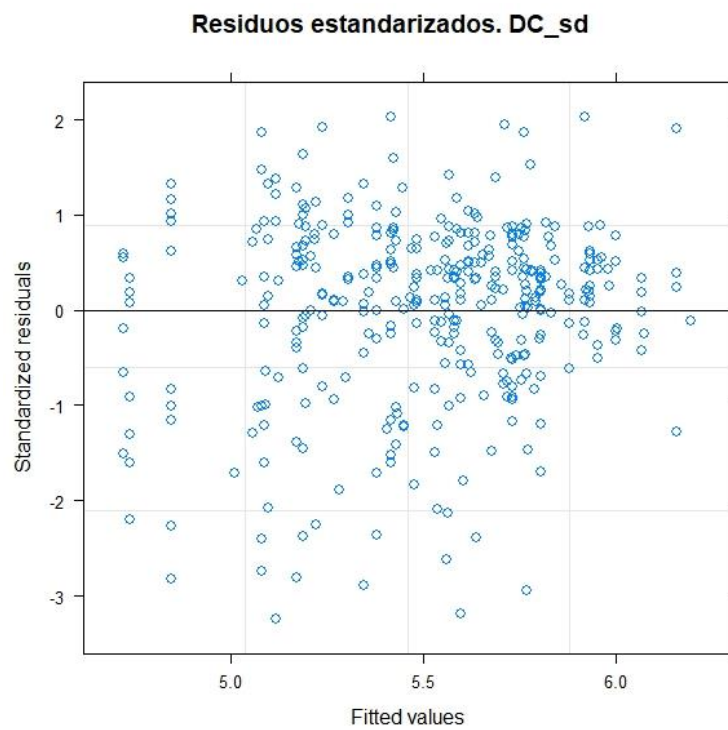
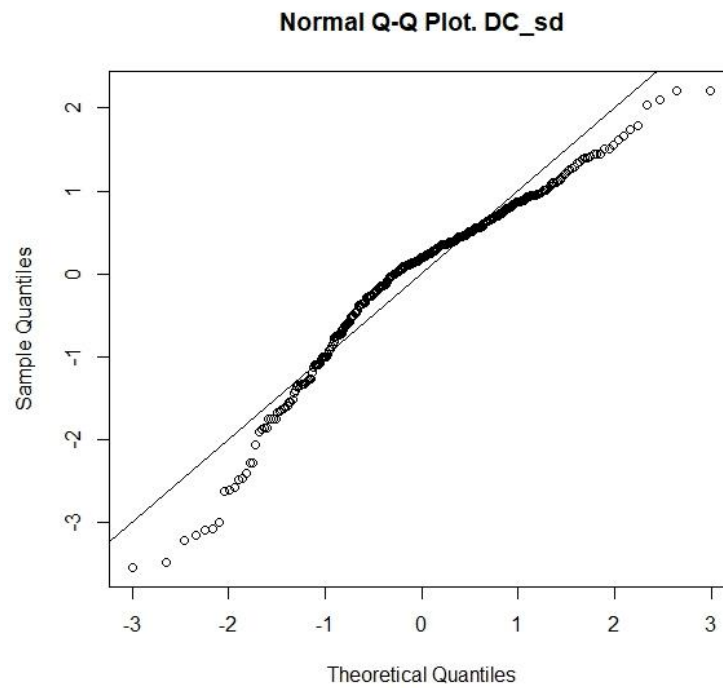
Number of Observations: 361
Number of Groups: 36

```

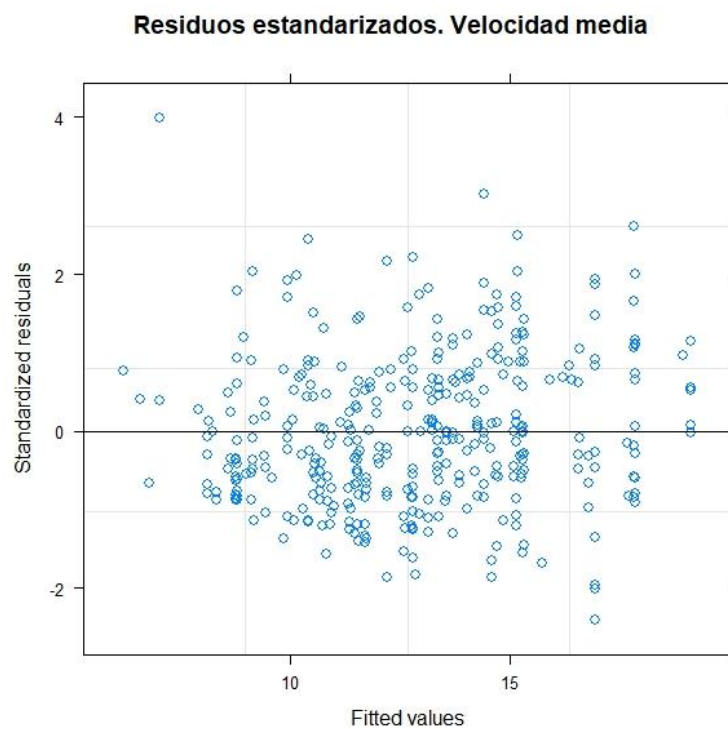
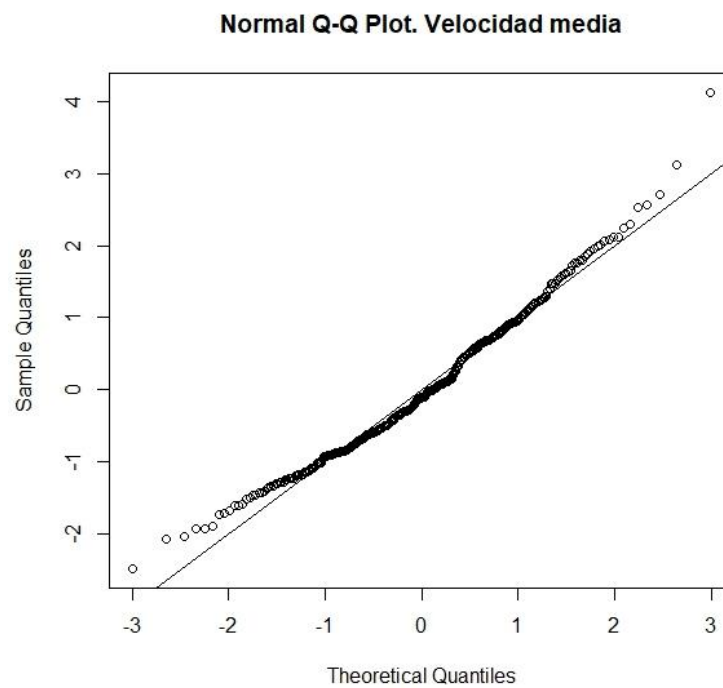
Validación DC m



Validación DC_sd



Validación media de la velocidad



```

> summary(m6sin_outlier)
Linear mixed-effects model fit by REML
Data: velocidad_df[-197, ]
      AIC      BIC    logLik
2317.372 2348.394 -1150.686

Random effects:
Formula: ~1 + tipo_dia | id
Structure: General positive-definite, Log-Cholesky parametrization
              StdDev   Corr
(Intercept)   1.566972 (Intr)
tipo_dialaborable 1.296135 0.858
Residual       5.617305

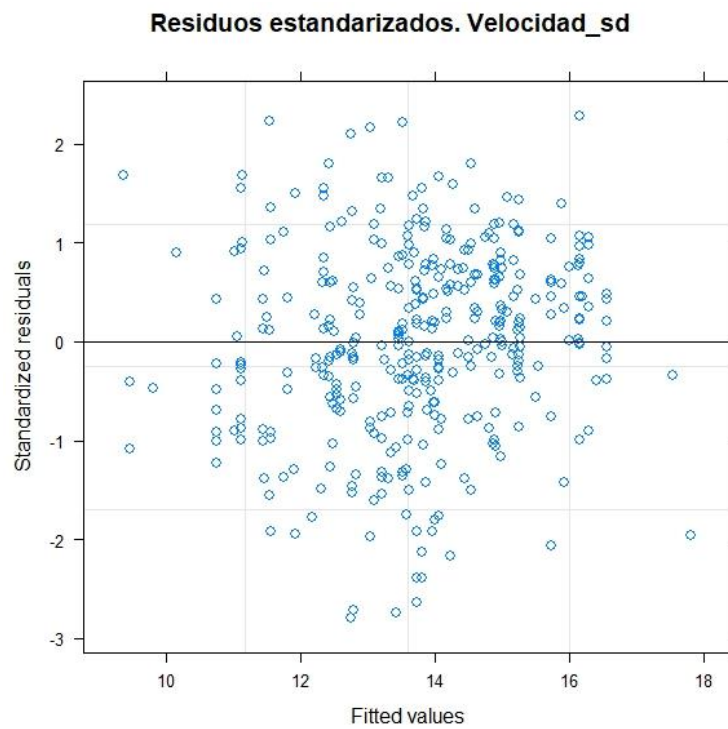
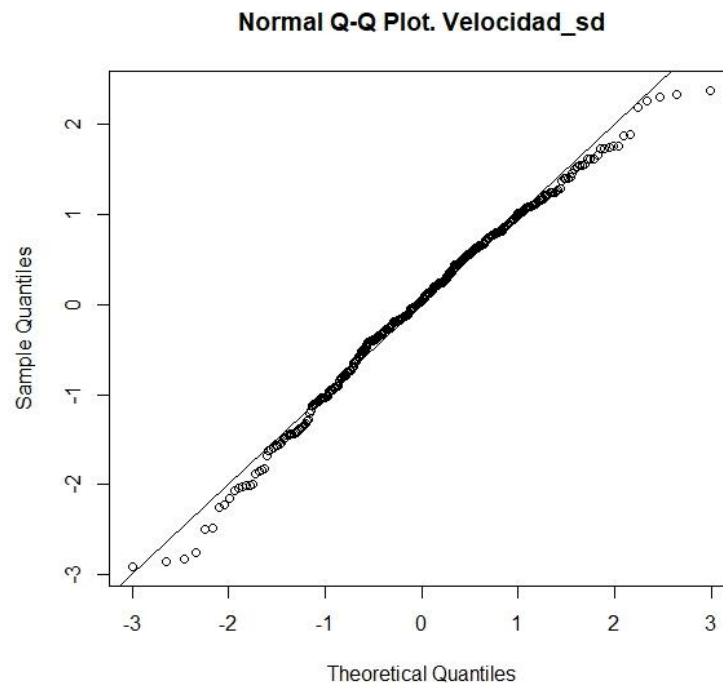
Fixed effects: speed_m ~ tipo_dia + momento
              Value Std.Error DF   t-value p-value
(Intercept)  10.316398 0.8177410 322  12.615727  0.0000
tipo_dialaborable  4.198149 0.8162190 322   5.143411  0.0000
momentotarde    -2.835804 0.7411490 322  -3.826227  0.0002
momentonoche    -0.142817 0.7894961 322  -0.180896  0.8566
Correlation:
              (Intr) tp_dlb mmnttr
tipo_dialaborable -0.627
momentotarde      -0.316 -0.110
momentonoche      -0.321 -0.083  0.437

Standardized Within-Group Residuals:
      Min       Q1       Med       Q3      Max
-2.4694022 -0.6934175 -0.1097386  0.6537373  3.1398981

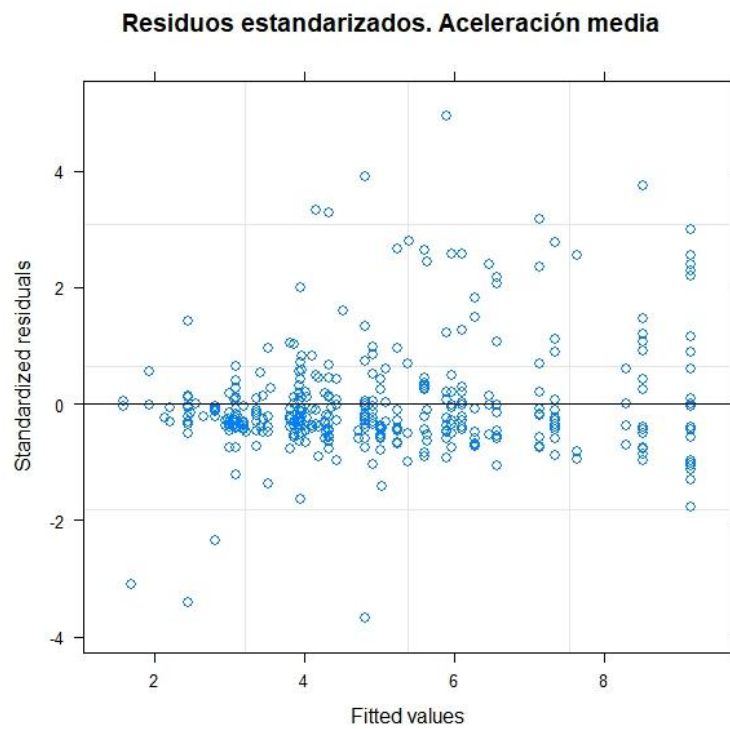
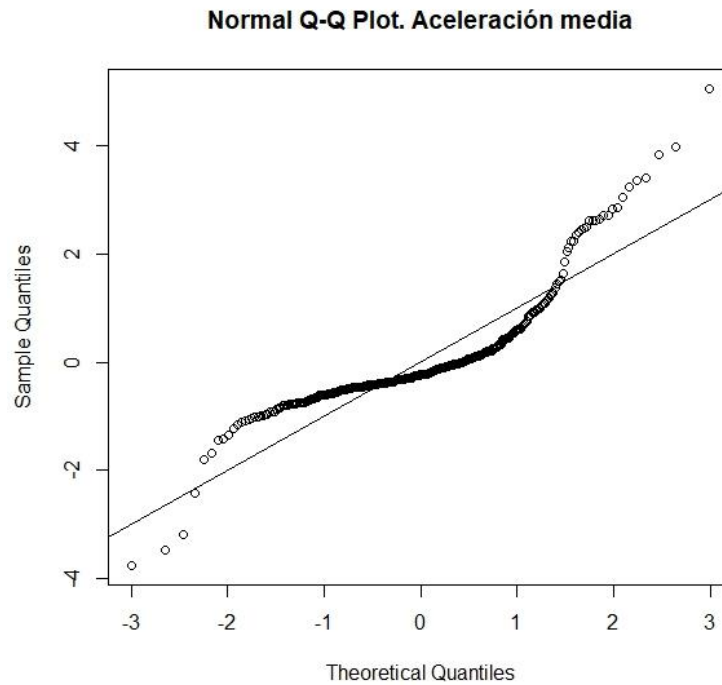
Number of Observations: 361
Number of Groups: 36

```

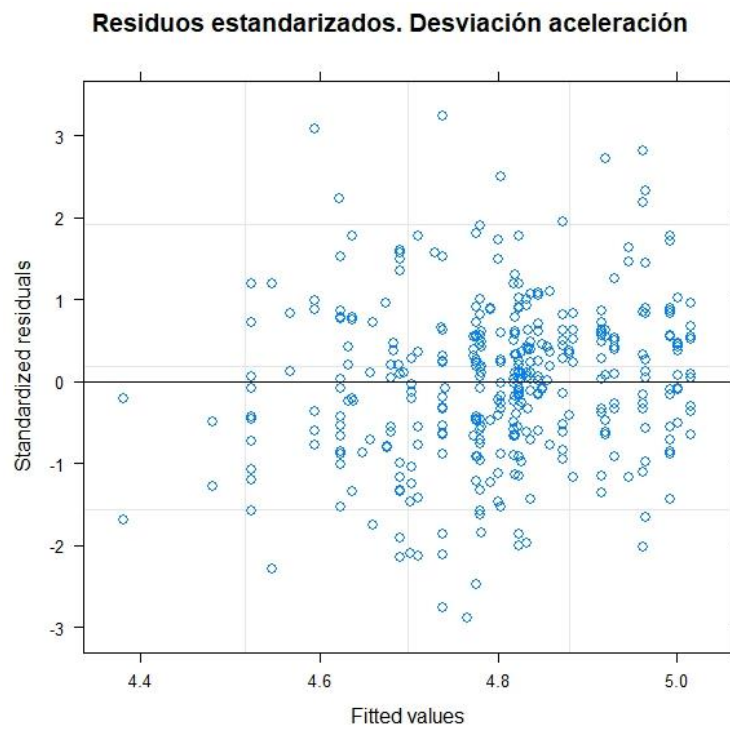
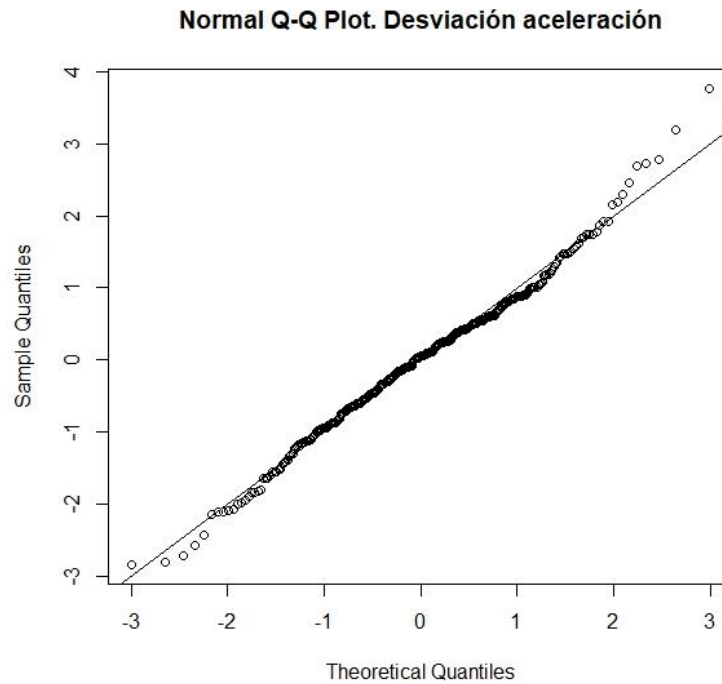
Validación desviaciones de la velocidad



Validación aceleración media



Validación desviaciones de la aceleración



Código R

```
install.packages("tidyverse")
install.packages("chron")
install.packages("trajr")
library(tidyverse)
library(readxl)
library(chron)
library(trajr)
library(openxlsx)
library(lubridate)
#setwd("C:/Users/aucoor/Desktop/TFM")

dades<-openxlsx::read.xlsx("C:/Users/aucoor/Documents/TFM/dades_audouin_jazel.xlsx",sheet=1)
#save.image("dades.RData")
dades<-as_tibble(dades)

#####
##### 1. PREPROCESSING DATOS -----
str(dades)
summary(dades)
# Corregim error de variable idViaje1
dades<-dades %>% mutate(idViaje1=ifelse(idViaje1=="V1b","V01",idViaje1))

# Pasamos a factor la variable DiaSem y ordenamos sus levels y recodificamos a numerico(?).
table(dades$DiaSem)
dades$DiaSem<-factor(dades$DiaSem,levels=c("lun","mar","mie","jue","vie","sab","dom"))
dades<- dades %>%
  mutate(diaset=recode(DiaSem,"lun"=1,"mar"=2,"mie"=3,"jue"=4,"vie"=5,"sab"=6,"dom"=7))
# Borramos DiaSem?

#Dia de la semana de cada fila
dades$dia1<- as.Date(dades$fechaOfici,origin="1899-12-30") #origen si excel
dades$weekday<-weekdays(dades$dia1)

# convertimos a hora la horaOficia
dades$horaOficia2<- chron(times=dades$horaOficia)
# Dado que hay trayectorias que duran >2 dias, hace falta saber la fecha y hora junta para calcular
la duracion
dades$fecha_completa<-as.POSIXct(with(dades, ymd(dia1) + hms(horaOficia2)))

#Cambiamos horaOficia a variable temps con el formato adecuado. Borramos horaOficia?
#dades$temps<-chron(times=dades$horaOficia)
```

```

#Dia des d'inici seguiment (duda:??)
#dades$dia<-1+(dades$fechaOfici-min(dades$fechaOfici))/86400
#dades$dia<-as.numeric(dades$dia)

#Calculamos tiempo por idViaje2: Para cada idViaje2 cogemos la horaOficia minima.
min.hora<-group_by(dades, idViaje2) %>%
  summarise(min_hora = min(fecha_completa, na.rm = TRUE))

#y añadimos esta nueva variable a los datos.
dades<-inner_join(dades,min.hora,by="idViaje2")

# Duda: que quiere decir times?
dades$times<-dades$fecha_completa-dades$min_hora
#dades$times<-hours(dades$)*3600+minutes(dades$times)*60+seconds(dades$times)

# Corregimos las variables binarias relacionadas con el lugar donde estan.
##### valores HABITAT: NA o rice_Field
##### valores HABITAT2: NA o on_land
dades$rice_field<-with(dades,ifelse(habitat==" ", "No", "Yes"))
dades$on_land<-with(dades,ifelse(habitat2==" ", "No", "Yes"))
dades$colony<-with(dades,ifelse(Colony==" ", "No", "Yes"))

# Creamos la PK: ANILLA + idViaje1
dades$PK<-paste0(dades$ANILLA,dades$idViaje1)
n_distinct(dades$PK) # 362 COMBINACIONES DIFERENTES

## 1.1 Creación variables ----
# Variables binarias: mañana , tarde, noche,

## Mañana: si vuela entre las 6: 13
dades$mati <- ifelse(between(hour(dades$fecha_completa),6,13 ), "si_mati", "no_mati")

## Tarde: si vuela entre las 14: 19
dades$tard <- ifelse(between(hour(dades$fecha_completa),14,19 ), "si_tard", "no_tard")

# Noche: si vuela entre las 20: 2
dades$nit <- ifelse(between(hour(dades$fecha_completa),20,24 ), "si_nit",
  ifelse(hour(dades$fecha_completa)==0, "si_nit",
    ifelse(between(hour(dades$fecha_completa),1,5 ), "si_nit", "no_nit")))

dades$rice_field <- factor(dades$rice_field)
dades$on_land <- factor(dades$on_land)
dades$colony <- factor(dades$colony)
dades$mati <- factor(dades$mati)
dades$tard <- factor(dades$tard)

```

```

dades$nit <- factor(dades$nit)

#1.2 Eliminación de variables ----
dades$campanya<-NULL # var constante
dades$ANILLA<-as.factor(dades$ANILLA)
dades$fechaOfici<-NULL # ya esta en fecha_completa ( y en dia1)
dades$horaOficia<-NULL # ya esta en fecha_completa ( y en horaOficia2)
dades$intervHora<-NULL # no util
dades$intervHo_1<-NULL # no util
dades$fechaHoraO<-NULL # casi siempre es igual que horaU excepto en los vuelos q son a la
madrugada...( de 0 a 1 o de 1 a 2 d la madrugada)
dades$fechaHoraU<-NULL # no util
dades$DiaSem<-NULL # ya esta en weekday
dades$diaset<-NULL # ya esta en weekday
dades$habitat<- NULL # ya esta en rice_field
dades$habitat2<-NULL # ya esta en on_land
dades$Colony<-NULL # ya esta en colony

#save(dades,file="datos_clean.RData")

# Comprobar que la 1a fila de cada PK , empieza en times=0
dades %>% group_by(PK) %>% summarise(tf=first(times)) %>% filter(tf!=0) %>% nrow()

# Ordenar dataframe segun times por grupo PK
#dades_ord<-dades %>% group_by(PK) %>% arrange(times,.by_group=TRUE)
#Comprobacion
#dades_ord %>% group_by(PK) %>% summarise(tf=first(times)) %>% filter(tf!=0) %>% nrow() # ok!
:)

#save(dades_ord,file="datos_clean_ord.RData")

#load("datos_clean_ord.RData")
# Creacion time_spent:

dades<-dades[order(dades$PK,dades$fecha_completa),]

dades$time_spent<-dades$times-lag(dades$times)
dades$time_spent<-with(dades,ifelse(time_spent<0,NA,time_spent))
#dades_net<-filter(dades,is.na(time_spent)==F)

summary(dades$time_spent)

dades %>%
  filter(between(time_spent,250,450))%>%
  ggplot(aes(x=time_spent/60)) +
  geom_density(fill="black",alpha=0.3)+

```



```

labs(title="Densidad del tiempo entre 2 localizaciones",x = "time_spent")+
theme_bw()

summary(dades$time_spent/60)

#####
#####2. CONSTRUCCIÓN INDICES -----
load("dades_clean_ord_vfinal.RData")

dades<-dades_ord
rm(dades_ord)

# nº de dias de cada trayectoria PK
n_dias<-dades %>% group_by(PK) %>% summarise(n=n_distinct(dia1)) %>% arrange(desc(n))
nrow(n_dias[n_dias$n!=1,])/nrow(n_dias) # 18,2% trayectorias tienen mas de 1 dia.(1 trayectoria
5107940V04 implica 4 dias!)

# Dia salida por cada trayectoria: se escoge el primer dia.
dias_salida<-dades %>% group_by(PK) %>% summarise(dia_salida =first(dia1),weekday_salida
=first(weekday))

# Objetivo script: Para cada una de las 362 combinaciones, calcular los indices necesarios para
caracterizar la trayectoria

#####
# CALCULO INDICES DE LAS TRAYECTORIAS
#####
# Solo longitud, latitud, fecha, anilla, idviaje1,idviaje2
coords <- data.frame(x = dades$LONG,y = dades$LAT,
                     times =
dades$times/60,ANILLA=dades$ANILLA,idViaje1=dades$idViaje1,idViaje2=dades$idViaje2,PK=dades$PK)

idxs<-c()
pk="5107940V04"
for (pk in unique(coords$PK)){
  coords_pk <- coords %>% filter(PK==pk)
  coords_pk2<-TrajFromCoords(coords_pk,timeCol="times") # x(long), y, time,
ANILLA,idviaje1,idviaje2,pk, displacementTime, polar, displacement
  smoothed_pk<- TrajSmoothSG(coords_pk2)
  derivs <- TrajDerivatives(smoothed_pk)
  speed_m <- mean(derivs$speed)
  speed_sd <- sd(derivs$speed)
  accel_m <- mean(derivs$acceleration)
  accel_sd <- sd(derivs$acceleration)
  # Turning angles

```

```

# Mitjana
r<- Mod(TrajMeanVectorOfTurningAngles(coords_pk2))
# SD
# If your trajectory does not have a constant step length, it should be _rediscretized_ by calling
TrajRediscretize before calling this function.
# Calculo r Y d=D/L: 1ro) discretizar para tener constant step length. 2do) Calcular r con la funcion
TrajStraightness
discret<- TrajRediscretize(coords_pk2,0.001)
#d<- TrajStraightness(discret)
#r<- Mod(TrajMeanVectorOfTurningAngles(discret)) en verdad es mejor rediscretizarla, pero no se
si al final hare servir R
sinuosidad<- TrajSinuosity2(discret)

## CALCULO DIRECTIONAL CHANGE.
DC_m <- mean(TrajDirectionalChange(coords_pk2))
DC_sd <- sd(TrajDirectionalChange(coords_pk2))
# RECOPIACION INDICES
idxs<-rbind(idxs,c(speed_m,speed_sd,accel_m,accel_sd,r,DC_m,DC_sd,sinuosidad))
}

idxs<- as.data.frame(idxs)
idxs$PK<- unique(coords$PK)
names(idxs)<-
c("speed_m","speed_sd","accel_m","accel_sd","r","DC_m","DC_sd","sinuosidad","PK")
idxs$dia_salida <- dias_salida$weekday_salida
idxs$tipo_dia<-ifelse(idxs$dia_salida%in% c("sábado","domingo"), "fin_de_semana"
,"dia_laborable")

hora_salida <-dades %>% group_by(PK) %>% summarise(first= first(fecha_completa),
last=last(fecha_completa))
hora_salida$duracion<- hora_salida$last- hora_salida$first # esta en formato minutos

# Añadimos al dataframe de los indices, estas nuevas variables (no hace falta leftjoin...)
idxs$duracion<-as.numeric(hora_salida$duracion)
idxs$hora_salida<- hora_salida$first
idxs$hora_final<- hora_salida$last

# Añadir a cada trayectoria: si sucede en la mañana, tarde, noche, madrugada

temporal<- dades %>% group_by(PK) %>%
summarise(mañana=any(mati=="si_mati"),tarde=any(tard=="si_tard"),noche=any(nit=="si_nit"))
idxs<- cbind(idxs,temporal)

#Añadimos los ultimos indices calculados.
#idxs$ebmax<-ebmaxs

```

```

#Borramos PK repetida
names(idxs)[15]<-"PK_1"
idxs$PK_1<-NULL
idxs<-idxs %>% select(PK,sinuosidad,duracion,everything())

idxs$id<-substr(idxs$PK,1,7)
idxs$tr<-substr(idxs$PK,8,10)
# Guardamos los indices
save(idxs,file="indices_vfinal.RData")

#####3.DESRIPTIVA PCA INDICES -----
#PCA ----
names(idxs2)
library(FactoMineR)

table(idxs2$bol)
idxs$dia_salida <-
factor(idxs$dia_salida,levels=c("lunes","martes","miércoles","jueves","viernes","sábado","domingo"),
ordered = T)
idxs$tipo_dia <- factor(idxs$tipo_dia,labels=c("fin_de_semana","laborable"))
idxs$momento <- factor(idxs$momento,levels=c("mañana","tarde","noche"),ordered = TRUE)

num_cols<-
c("sinuosidad","duracion","speed_m","speed_sd","accel_m","accel_sd","DC_m","DC_sd")
cat_cols <-c("dia_salida","momento")
pca1<-PCA(idxs,c(num_cols,cat_cols)), ncp=5,scale.unit = T,quali.sup=c(9:10))# la definitiva 20/08
fviz_pca_var(pca1, col.var = "black")

### eigenvalues
plot(pca1$eig[, "eigenvalue"],type = "p")
lines(pca1$eig[, "eigenvalue"])

# variabilidad explicada
barplot(pca1$eig[, "cumulative percentage of variance"])

##3.1. Representacion de las trayectorias en el plano factorial----

pca_df<-as.data.frame(pca1$ind$coord)
pca_df$momento<-idxs$momento
pca_df$tipo_dia<-idxs$tipo_dia
pca_df$dia_salida<-idxs$dia_salida
pca_df %>%
  ggplot(aes(x=Dim.1,y=Dim.2))+
  geom_point()+
  geom_hline(yintercept=0,linetype="dashed",color="blue")+

```

```
geom_vline(xintercept=0, linetype="dashed",color="blue")+
theme_bw()+
labs(title="Representación de las trayectorias en el plano factorial")
#ggsave("C:/Users/auacor/Documents/TFM/pca1.jpeg",width = 7,height = 5)
```

#3.2. Representacion de los individuos en el plano factorial segun una variable categorica----

```
#tipo_dia
pca_df %>%
  ggplot(aes(x=Dim.1,y=Dim.2,color=tipo_dia))+
  geom_point()+
  geom_hline(yintercept=0,linetype="dashed",color="blue")+
  geom_vline(xintercept=0, linetype="dashed",color="blue")+
  theme_bw()+
  labs(title="Representación de las trayectorias en el plano factorial")
#momento
pca_df %>%
  ggplot(aes(x=Dim.1,y=Dim.2,color=momento))+
  geom_point()+
  geom_hline(yintercept=0,linetype="dashed",color="blue")+
  geom_vline(xintercept=0, linetype="dashed",color="blue")+
  theme_bw()+
  labs(title="Representación de las trayectorias en el plano factorial")
```

PLOT OF CATEGORICAL VARIABLES. Ver solo los centros de gravedad de las cualitativas(que son suplementarias)

```
plot(pca1,invisible=c("ind"),cex=0.8)
```

##3.4. Representacion CONJUNTA----

```
#fm = round(max(abs(Psi[,1])))
fm=0.87
iden = row.names(idxs)
etiq = names(idxs[,c(2:7,9:10)])
ze = rep(0,length(etiq))
U = pca1$var$coord
plot(pca1$ind$coord[,1],pca1$ind$coord[,2],col="gray",xlim=c(-1,1.2),ylim=c(-0.9,1),xlab = "Dim 1
(49.28%)",ylab="Dim 2 (14.49%)",
  main="Representación conjunta: índices con las variables suplementarias")
arrows(ze, ze, fm*U[,1], fm*U[,2], length = 0.07,col="black")
text(fm*U[,1],fm*U[,2],labels=etiq,col="black", cex=1)
dcat<-c(11,23)
colors<-c("red","darkgreen")
c<-1
for(k in dcat){
  sequestColor<-colors[c]
  print(sequestColor)
```

```

fdic1 = tapply(pca1$ind$coord[,1],idxs[,k],mean)
fdic2 = tapply(pca1$ind$coord[,2],idxs[,k],mean)
# points(fdic1,fdic2,pch=16,col=seguentColor, labels=levels(dd[,k]))
text(fdic1,fdic2,labels=levels(idxs[,k]),col=seguentColor, cex=1)
c<-c+1
}
legend("topleft",names(idxs[,dcat]),col=colors,cex=1,pch=20)

```

##3.5. Contribucion variables ----

```

var <- get_pca_var(pca1)
contributions <- var$contrib
View(contributions)
corrplot(var$contrib[,1:2], is.corr=FALSE)
fviz_contrib(pca1, choice = "var", axes = 1, top = 5,title="Contribución de los índices en la 1a
componente principal")
ggsave("C:/Users/aucoor/Documents/TFM/pca_contribucion1.jpeg",width =7 ,height =6 )
fviz_contrib(pca1, choice = "var", axes = 2, top = 5,title="Contribución de los índices en la 2a
componente principal")
ggsave("C:/Users/aucoor/Documents/TFM/pca_contribucion2.jpeg",width =7 ,height =6 )
res.desc <- dimdesc(pca1, axes = c(1,2), proba = 0.05)
res.desc$Dim.1
res.desc$Dim.2

```

#####4. MODELOS MIXTOS Y REPETIBILIDAD

#Se adjunta en el trabajo, solo el código referente a el índice de sinuosidad

```

library(tidyverse)
library(nlme)
# Función para calcular el ICC ----

```

```

icc=function(dataset,y,ind){
  require(nlme)
  model.lme=lme(y~1,data=dataset,random=~1|ind) # ajusta modelo solo con efecto aleatorio de
gaviota. No hay covariables
  tau<-(unlist(model.lme$modelStruct$reStruct))+log(model.lme$sigma) # no lo entiendo
  tau.res<-log(model.lme$sigma) # logaritmo de la std del error aleatorio
  sa<-exp(2*tau) # exponencial de 2* tau
  se<-(model.lme$sigma^2) # se= varianza del error aleatorio
  icc=sa/(sa+se)

```

```

#Standard error
s.tau<-(model.lme$apVar)[1]          ## es la matriz sigma pero en los logs de la varianza...
hemos d pasarlo a la escala
s.res<-(model.lme$apVar)[4]
s.tau.res<-(model.lme$apVar)[2]
var.sa<-4*exp(4*tau)*s.tau

```

```

var.se<-4*exp(4*tau.res)*s.res
cov.sa.se<-4*exp(2*(tau+tau.res))*s.tau.res

sigma<-array(c(2,2))
sigma[1,]<-c(var.sa,cov.sa.se)
sigma[2,]<-c(cov.sa.se,var.se)
dev.sa<-(1-icc)/(sa+se) # son derivadas hechs a mano. derivada de ICC resepecto cada
ocmponente q hay en el ICC
dev.se=(-1)*icc/(sa+se)
dev<-c(dev.sa,dev.se)
seicc<-t(dev)%*%sigma%*%dev
z<-0.5*log((1+icc)/(1-icc))
sez<-seicc/(((1+icc)**2)*((1-icc)**2))
llz<-z-1.96*sqrt(sez)
ulz<-z+1.96*sqrt(sez)
ll95 = (exp(2*llz)-1)/(exp(2*llz)+1);
ul95 = (exp(2*ulz)-1)/(exp(2*ulz)+1);
#Output
est<-round(c(icc,sqrt(seicc),ll95,ul95,z),digits=4)
names(est)<-c("ICC", "SE ICC", "LL CI95%", "UL CI95%", "Z")
return(list(model.lme,est))
}

```

```
icc2 <- possibly(icc, otherwise = NA)
```

```
library(nlme)
```

```

# Variable momento del día
momentos <- dades %>%
  group_by(PK) %>%
  summarise(m=first(mati),t=first(tard),n=first(nit))

```

```

momentos$momento <- with(momentos,ifelse(m=="si_mati","mañana",
  ifelse(t=="si_tard","tarde","noche")))

```

```

idxs<-left_join(idxs,momentos)
idxs$momento<-factor(idxs$momento, levels=c("mañana","tarde","noche"))
table(idxs$momento)

```

```
# 4.1. SIN U O S I D A D -----
```

```

idxs %>% ggplot(aes(sinuosidad))+
  geom_density(fill="black",alpha=0.4)+
  theme_bw()+
  labs(title="Distribución de la sinuosidad")
#ggsave("C:/Users/aucor/Documents/TFM/dist_sinuosidad.jpeg",width =4 ,height =5 )

```

```

idxs %>% ggplot(aes(log(sinuosidad)))+
  geom_density(fill="black",alpha=0.4)+
  theme_bw()+
  labs(title="Distribución del logaritmo de la sinuosidad")
#ggsave("C:/Users/auco/Documents/TFM/dist_sinuosidad_log.jpeg",width =4 ,height =5 )

# Se usa el logaritmo de la sinuosidad.
# 1.1 Definir la mejor estructura de varianzas y covarianzas. ----

# Modelo 1: Efecto aleatorio homoscedastico Y ERROR RESIDUAL HOMOSCEDATICO
m1<-lme(log(sinuosidad)~1, random = ~1|id,data=idxs,method="ML")

# Modelo 2: Efecto aleatorio homoscedastico Y varianza residual diferente segun el tipo de día.
m2<-lme(log(sinuosidad)~1, random =
~1|id,weights=varIdent(form=~1|tipo_dia),data=idxs,method="ML")

# Modelo 3: Varianza entre gaviotas diferente segun tipo dia Y ERROR RESIDUAL
HOMOSCEDATICO
m3<-lme(log(sinuosidad)~1, random = ~1+tipo_dia|id,data=idxs,method="ML")

#Modelo 4: todo heteroscedastico

m4<-lme(log(sinuosidad)~1, random =
~1+tipo_dia|id,weights=varIdent(form=~1|tipo_dia),data=idxs,method = "ML")

AIC(m1)
AIC(m2)
AIC(m3)
AIC(m4)

anova(m1,m2) # gana m1 (pvalor>0.05)
anova(m1,m3) # gana m3
anova(m3,m4) #gana m3 , pero x poco--> s usa AIC

## 1.2 Introducción efecto tipo_dia ----
ctrl <- lmeControl(opt='optim')
m5 <-lme(log(sinuosidad)~tipo_dia, random =
~1+tipo_dia|id,weights=varIdent(form=~1|tipo_dia),data=idxs,method="ML",control=ctrl)
m6 <-lme(log(sinuosidad)~tipo_dia+momento, random =
~1+tipo_dia|id,weights=varIdent(form=~1|tipo_dia),data=idxs,method="ML",control=ctrl)
m7 <-lme(log(sinuosidad)~tipo_dia*momento, random =
~1+tipo_dia|id,weights=varIdent(form=~1|tipo_dia),data=idxs,method="ML",control=ctrl)

AIC(m5)
AIC(m6)

```

AIC(m7)

```
anova(m5,m6) # gana m6 (pvalor<0.05)
```

```
anova(m6,m7) # gana m6 (pvalor>0.05)
```

```
idxs%>% group_by(tipo_dia) %>%  
  ggplot(aes(x=tipo_dia,y=sinuosidad,fill=tipo_dia))+  
  geom_boxplot() +  
  facet_wrap(~momento)+  
  theme_bw()+  
  guides(fill=FALSE)+  
  labs(title="Distribución de la sinuosidad según tipo de día y momento del día")  
ggsave("C:/Users/aucoor/Documents/TFM/sinuosidad.jpeg",width =7 ,height =6 )
```

```
idxs %>% group_by(tipo_dia) %>%  
  ggplot(aes(x=tipo_dia,y=sinuosidad,fill=tipo_dia))+  
  geom_boxplot() +  
  theme_bw()+  
  guides(fill=FALSE)+  
  labs(title="Según tipo de día")  
ggsave("C:/Users/aucoor/Documents/TFM/sinuosidad_dia.jpeg",width =4 ,height =5 )
```

```
idxs %>% group_by(momento) %>%  
  ggplot(aes(x=momento,y=sinuosidad,fill=momento))+  
  geom_boxplot() +  
  theme_bw()+  
  guides(fill=FALSE)+  
  labs(title="Según el momento del día")  
ggsave("C:/Users/aucoor/Documents/TFM/sinuosidad_diaaaa.jpeg",width =4 ,height =5 )
```

1.3 Summary mejor modelo y diagnostico ----

```
idxs$tipo_dia<-factor(idxs$tipo_dia,levels=c("festivo","working_day"))
```

```
levels(idxs$tipo_dia)<-c("fin_de_semana","laborable")
```

```
idxs$tipo_dia = factor(idxs$tipo_dia,levels(idxs$tipo_dia)[c(2,1)])
```

```
m6 <- lme(log(sinuosidad)~tipo_dia+momento, random =
```

```
~1+tipo_dia|id,weights=varIdent(form=~1|tipo_dia),data=idxs,method="REML",control=ctrl)
```

```
summary(m6)
```

```
anova(m6)
```

```
intervals(m6)
```

```
# Diagnostico modelo m0: errores bastantes normales
```

```
qqnorm(scale(resid(m6)),main="Normal Q-Q Plot. Log sinuosidad")
```

```
abline(0,1)
```

```
plot(m6,main="Residuos estandarizados. Log sinuosidad") # decir que los residuos estan entorno al 0.
```

```
plot( m6, resid(.) ~ fitted(.) | tipo_dia, id = .02)
```



```
## 1.3a Conflicto interaccion tarde/no tarde ----
```

```
tarde
m6b <- lme(log(sinuosidad)~tipo_dia*tarde, random =
~1+tipo_dia|id,weights=varIdent(form=~1|tipo_dia),data=idxs,method="REML",control=ctrl)
anova(m6b)
```

```
## 1.4 lcc ----
```

```
idxs$y<-log(idxs$sinuosidad)
idxs$ind<-idxs$id
icc(idxs,y,id)
```

```
#1r) Modelo por separado para cada tipo de dia--> 2 indices
# Calculo la diferencia de medias en los datos originales sin permutar
icc_w<-icc(idxs[idxs$tipo_dia=="laborable"],y,id)
icc_f<-icc(idxs[idxs$tipo_dia=="fin_de_semana"],y,id)
dif_original<- icc_w[[2]][1]-icc_f[[2]][1] #-0.2718
```

```
#Permutamos y calculamos la diferencia de ICC's
nperm = 1000
set.seed(2802)
f.perms = replicate(nperm, {
  dades.perm = unlist(tapply(idxs$tipo_dia, idxs$ind, sample, replace = FALSE)) # sin
reemplazamiento para mantener la estructura del diseño
  ICC_work <-icc2(idxs[dades.perm=="working_day"],y,ind)
  ICC_fest <-icc2(idxs[dades.perm=="festivo"],y,ind)
  if (NA %in% ICC_work) {
    ICC_work<-NA
  }else{
    ICC_work<-ICC_work[[2]][1]
  }
  if (NA %in% ICC_fest) {
    ICC_fest<-NA
  }else{
    ICC_fest<-ICC_fest[[2]][1]
  }
  difs<-ICC_work-ICC_fest # calculo en cada permutacion, las diferencias
}
)
f.perms<- f.perms[-which(f.perms %in% NA)]
```

```
# Conclusion: pvalor= % de veces en que dif permutada es mayor a la dif original (ambas en
v.absoluto)
sum(abs(f.perms) > abs(dif_original))/nperm # two-tailed test
#pvalor>0.05--> No reb Ho de igualdad de medias de ICC.
```

5. SHINY APP

```
library(shiny)
library(shinydashboard)
#library(tidyverse)
library(dplyr)
#library(ggplot2)
library(leaflet)
library(trajr)
#library(broom)
library(imager)
setwd("C:/Users/aucor/Documents/TFM")

aux<-function(coords1,trip,i) {
  data<-coords1 %>% filter(idViaje1==trip[i])
  TrajFromCoords(data,timeCol="times")
}
# dades$mati<-factor(dades$mati)
# dades$tard<-factor(dades$tard)
# dades$nit<-factor(dades$nit)
# temps_traj<-dades %>% group_by(PK) %>%
summarise(m=table(mati)[2],n=table(nit)[2],t=table(tard)[2],total=n())
# temps_traj$total2<- rowSums(temps_traj[,2:4],na.rm=T)
# temps_traj$comprov<-ifelse(temps_traj$total2==temps_traj$total,T,F)
# # % de tiempo para cada trayectoria
# temps_traj$perc_mati<-round(100*temps_traj$m/temps_traj$total,1)
# temps_traj$perc_tard<-round(100*temps_traj$t/temps_traj$total,1)
# temps_traj$perc_nit<-round(100*temps_traj$n/temps_traj$total,1)
# save(temps_traj,file = "temps_traj.RData")

# Carga de los archivos necesarios para la app
load("dades_clean_ord_vfinal.RData")
load("temps_traj.RData",verbose = T)
load("indices_vfinal.RData",verbose = T)
idxs$sinuosidad <-round(idxs$sinuosidad,2)
idxs$DC_m <-round(idxs$DC_m,2)

coords <- data.frame(x = dades$LONG,y = dades$LAT,
  times = dades$times,ANILLA=dades$ANILLA,
  idViaje1=dades$idViaje1,idViaje2=dades$idViaje2,PK=dades$PK,
```

```

    mati=dades$mati,tard=dades$tard,nit=dades$nit,
    fecha_completa=dades$fecha_completa)

# Funcion colores para pintar cada linea de 2 puntos segun si es mañana, tarde,noche
colores<-function(df){
  color<-c()
  for( i in 1: nrow(df)){
    if (df$mati[i]=="si_mati") {
      color<-c(color,"green")
    }else if(df$tard[i]=="si_tard"){
      color<-c(color,"orange")
    }else if(df$nit[i]=="si_nit"){
      color<-c(color,"blue")
    }else {
      color<-c(color,"black")
    }
  }
  return(color)
}

# Barra de pestaña
sidebar<- dashboardSidebar(
  sidebarMenu(
    menuItem("Trayectorias-Visualización", tabName = "TrayectoriasVisualización", icon =
icon("feather") ),
    menuItem("Trayectorias-Ranking", tabName="ranking",icon=icon("feather"))

  )
)

body<-dashboardBody(
  tags$style(type = "text/css", "#map {height: calc(100vh - 80px) !important;}"),
  tabItems(
    menuItem(tabName = "TrayectoriasVisualización",
      strong(h3("Elige una trayectoria para más información")),
      #LAYOUT DE LA PESTAÑA: panel sidebar + panel principal
      # PANEL SIDEBAR
      fluidPage(
        fluidRow(
          column(width=3,selectInput("idgaviota", "Selecciona ID gaviota",choices = unique(idxs$Id),
            selected = unique(idxs$Id[1]) )),
          column(width=3,htmlOutput("idviaje"))
        ),

        # PANEL PRINCIPAL
        h3(strong(textOutput("titulo1"))),

```

```

column(width=12,leafletOutput("mapa1",height=600)),

valueBoxOutput("manyana",width=2),
valueBoxOutput("tarde",width=2),
valueBoxOutput("noche",width=2)
#fluidRow(h3(strong("Inicio: "),textOutput("inicio"))),
#fluidRow(h3(strong("Final : "),textOutput("final"))),
)
),# FIN PESTAÑA 1

tabItem(tabName = "ranking",
  fluidPage(
    fluidRow(
      column(width=3,radioButtons("indice", "Selecciona un índice:",choices =
c("sinuosidad","duracion","DC_m"),
        selected = "sinuosidad" )),
      column(width=3,radioButtons("TOP", "TOP:",choices = c("3" ,"6","10"),
        selected = "3" )),
      column(width=3,radioButtons("manera","De manera:",choices =
c("ascendente","descendente"),
        selected = "ascendente"))
    ),
    column(width=12,plotOutput("maparanking",height=700))
    #FIN PESTAÑA 2
  )
)
)
)

##### UI #####
# Una vez definido el sidebar y el body--> UI
ui <- dashboardPage(
  dashboardHeader(
    title = "Visualización trayectorias",titleWidth = 600
  ), sidebar,
  body
)

##### SERVER #####

server <- function(input, output) {
output$idviaje<-renderUI({
  idviajes<-idxs %>% filter(id==input$idgaviota) %>% select(tr)
  selectInput(inputId = "idviaje2",label = "Selecciona un ID viaje",choices=unique(idviajes$tr))
})
output$idviaje_ranking<-renderUI({

```

```

idviajes<-idxs %>% filter(id==input$idgaviota) %>% select(tr)
selectInput(inputId = "idviaje2_ranking",label = "Selecciona un ID
viaje",choices=unique(idviajes$tr))
})
output$manyana<- renderValueBox({
  valor<-temps_traj %>% filter(PK==paste0(input$idgaviota,input$idviaje2))
  valueBox(
    paste0(valor$perc_mati,"%"), "tiempo mañana", icon = icon("sun"),
    color = "green"
  )
})
output$tarde<- renderValueBox({
  valor<-temps_traj %>% filter(PK==paste0(input$idgaviota,input$idviaje2))
  valueBox(
    paste0(valor$perc_tard,"%"), "tiempo tarde", icon = icon("sun"),
    color = "orange"
  )
})
output$noche<- renderValueBox({
  valor<-temps_traj %>% filter(PK==paste0(input$idgaviota,input$idviaje2))
  valueBox(
    paste0(valor$perc_nit,"%"), "tiempo noche", icon = icon("sun"),
    color = "blue"
  )
})
output$titulo1<- renderText({
  paste0("Trayectoria id gaviota= ",input$idgaviota,". Viaje ",input$idviaje2)
})
output$mapa1<- renderLeaflet({
  # Filtramos todos los viajes de una determinada gaviota
  coords1<-coords %>% filter(ANILLA == input$idgaviota)
  trips<-unique(coords1$idViaje1)
  trj<-NULL
  # Calculo de los limites del mapa de cada gaviota
  for (i in 1:length(trips)){
    trj[[i]]<-aux(coords1,trips,i)
  }
  x1<-min(sapply(1:length(trips), function(i) min(trj[[i]]$x)))
  x2<-max(sapply(1:length(trips), function(i) max(trj[[i]]$x)))
  y1<-min(sapply(1:length(trips), function(i) min(trj[[i]]$y)))
  y2<-max(sapply(1:length(trips), function(i) max(trj[[i]]$y)))

  #Para cada gaviota, plot cada viaje
  # title <- tags$div(

```

```

# tag.map.title, HTML(paste0("Trayectoria id gaviota= ",input$idgaviota,". Viaje
",input$idviaje2))
# )
#filtramos la trayectoria
coords2<-coords1 %>% filter(idViaje1==input$idviaje2)
# extraer info de esa trayectoria--> idxs2
traject<-idxs %>% filter(PK==paste0(input$idgaviota,input$idviaje2))
inicio_x<-coords2$x[1]
inicio_y<-coords2$y[1]
final_x<-coords2$x[nrow(coords2)]
final_y<-coords2$y[nrow(coords2)]
m<-leaflet(coords2) %>%
  addTiles() %>%
  #fitBounds(lng1 = x1, lat1 = y1, lng2 = x2, lat2 = y2 ) %>%
  addPolylines(lng = ~x, lat = ~y,color="grey") %>%
  addCircleMarkers(lng = ~x, lat = ~y,radius = 3,color=colores(coords2)) %>%
  addMarkers(lng = inicio_x, lat = inicio_y,
    popup = paste(strong("Inicio:"), first(coords2$fecha_completa), "<br>",
      strong("Final:"), last(coords2$fecha_completa), "<br>",
      strong("Duración: "),paste0(traject$duracion %/% 60," horas y ",
round(traject$duracion %%%60,2), " minutos") , "<br>",
      strong("Sinuosidad: "),round(traject$sinuosidad,2), "<br>",
      strong("Día de salida: "),traject$dia_salida
    )
  ) #)%>%
  #addControl(title,className="map-title") # en addContorl s puede especificar lng1, lat1, lng2,
lat2
  #latitudes and longitudes of the south-west and north-east corners of rectangles
})

output$maparanking <-renderPlot({

  if (input$manera=="ascendente") {
    #imgs<-idxs %>% arrange(idxs[,input$indice]) %>% top_n(as.numeric(input$TOP),idxs[,
input$indice])
    imgs<-idxs %>% arrange(!as.name(input$indice)) %>% top_n(as.numeric(input$TOP),-
!!as.name(input$indice))

  }else{
    imgs<-idxs %>% arrange(!as.name(input$indice)) %>%
top_n(as.numeric(input$TOP),!!as.name(input$indice))
  }
  files <- paste0(imgs$id,"_viaje",imgs$tr)
  files2 <- paste0(getwd(),"trajs/",files,".jpeg")
  list_of_images = lapply(files2, load.image)
  if (input$TOP=="3") {

```

```

layout(matrix(1:3,nr=1,byr=T))
if (input$indice=="duracion") {
  for (j in 1:3) plot(list_of_images[[j]],axes = FALSE,main=paste0(input$indice," ",
                                                                    imgs[j,input$indice] %/% 60, " horas y ",
                                                                    round(imgs[j,input$indice] %/%60,2), " minutos")
                                                                    )
}
}else{
  for (j in 1:3) plot(list_of_images[[j]],axes = FALSE,main=paste0(input$indice," ",
                                                                    imgs[j,input$indice]))
}
}else if (input$TOP=="6") {
  layout(matrix(1:6,nr=2,byr=T))
  if (input$indice=="duracion") {
    for (j in 1:6) plot(list_of_images[[j]],axes = FALSE,main=paste0(input$indice," ",
                                                                    imgs[j,input$indice] %/% 60, " horas y ",
                                                                    round(imgs[j,input$indice] %/%60,2), " minutos")
                                                                    )
  }
}else{
  for (j in 1:6) plot(list_of_images[[j]],axes = FALSE,main=paste0(input$indice," ",
                                                                    imgs[j,input$indice]))
}
}else{
  layout(matrix(1:10,nrow=2,ncol=5,byr=T))
  if (input$indice=="duracion") {
    for (j in 1:10) plot(list_of_images[[j]],axes = FALSE,main=paste0(input$indice," ",
                                                                    imgs[j,input$indice] %/% 60, " horas y ",
                                                                    round(imgs[j,input$indice] %/%60,2), " minutos")
                                                                    )
  }
}else{
  for (j in 1:10) plot(list_of_images[[j]],axes = FALSE,main=paste0(input$indice," ",
                                                                    imgs[j,input$indice]))
}
}
})
}

```

```

# Run the application
shinyApp(ui = ui, server = server)

```